Copyright

by

Elizabeth Rose Milano

2015

The Dissertation Committee for Elizabeth Rose Milano Certifies that this is the approved version of the following dissertation:

The Genetic Architecture of Quantitative Traits in Locally Adapted Plant Ecotypes

Committee:

Thomas E. Juenger, Supervisor

Mark Kirkpatrick

Craig Randal Linder

Alan Lloyd

Noland Martin

The Genetic Architecture of Quantitative Traits in Locally Adapted Plant Ecotypes

by

Elizabeth Rose Milano, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2015

Acknowledgements

This work is a product of the time and support of so many people that I have had the opportunity to interact with during my tenure as a graduate student. I would first like to acknowledge my supervisor, Dr. Tom Juenger, without which none of this would be possible. Tom provided opportunities and guidance that allowed me to explore the wide world of quantitative genetics and evolutionary ecology. I would also like to acknowledge the collaborators for each of these projects: Dr. Amanda Kenny (Chapter 1), Dr. David Lowry (Chapter 2), and Courtney Payne (Chapter 3). The Juenger Lab has grown exponentially over the past several years and I have benefitted from collaborating with a diverse set of talented postdocs and technicians, notably Dr. John Lovell, Dr. Kyle Hernandez, and Jason Bonnette. Special thanks to Dr. Dave Des Marais for guidance and witty office banter during my early impressionable years.

I would also like to acknowledge the truly Integrative Biology Department at UT. Special thanks to Dr. Robin Hopkins for her floral know-how and Dr. Jesse Weber for his RADical genotyping skills. Thanks to all the brilliant friends I've made here, especially those who forged the trail ahead, Dr. Emily Jane McTavish, Dr. Geneviève Smith, Dr. Samuel Scarpino, Dr. Ginnie Morrison, Dr. Roz Eggo, Dr. Tinisha Hancock, Dr. Gayle Burstein, Rebecca Tarvin, Mariska Brady, Catalina Cuellar, and so many others. Thanks as well to Kelsey Hopson for support and shelter during the summer months in Colorado.

Lastly, I would like to thank my family. From a young age my parents have facilitated naturalist and scientific endeavors through camping trips and science fair projects and I am eternally grateful for all the experiences that have shaped my curiosity of the natural world.

The Genetic Architecture of Quantitative Traits in Locally Adapted Plant Ecotypes

Elizabeth Rose Milano, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Thomas E. Juenger

Locally adapted ecotypes are a common phenomenon generating plant diversity within species, yet we know surprisingly little about the genetic mechanisms that lead to locally adapted traits. The genetic architecture underlying traits can indicate evolutionary history and predict response to selection, with applications in evolutionary ecology, conservation, and crop development. This research broadly investigates the genetic architecture of quantitative traits in paired ecotypes from different plant species. I used multivariate comparative methods and quantitative trait loci (QTL) mapping to quantify genetic correlations and population divergence, between ecologically relevant traits, both at the phenotypic and genotypic level. I tested for adaptive floral trait evolution in a perennial wildflower by comparing differentiation at neutral loci to differentiation in a suite of quantitative floral traits in an *Ipomopsis aggregata* hybrid zone. I used multivariate comparisons to incorporate the genetic covariance architecture underlying floral display and reward traits, and found a strong signal for divergent selection. Non-neutral divergence for multivariate quantitative traits suggests that selection by pollinators is maintaining a correlation between floral display and reward. In *Panicum virgatum*, a native perennial grass, I used a genetic mapping population, segregating ecotypic variation, to construct a linkage map, and map QTL for nine ecological traits. Most QTL had intermediate to small effects and clustered on a limited number of linkage

groups. I also found over half of the functional allelic effects displayed patterns associated with fixed differences between ecotypes. These results suggest there is considerable standing genetic variation within local populations, as well as between ecotypes for ecologically relevant traits. Lastly, I explored the genetics of plant tissue quality in *Panicum hallii*, a model lignocellulosic grass system. Cell wall components compose the bulk of lignocellulosic biomass and contribute to the recalcitrance of plant tissue. I characterized the divergence of four major cell wall components between ecotypes, identified 14 QTL, and found half of the QTL localized to a single linkage group. Exploring the genetic architecture of tissue traits in a tractable system will lead to a better understanding of cell wall structure and function as well as provide genomic resources for bioenergy crop improvement.

Table of Contents

List of Tables

List of Figures

Chapter 1: Adaptive differentiation in floral traits in the presence of high gene flow in scarlet gilia (*Ipomopsis aggregata*)

Abstract

Plant-pollinator interactions are thought to be major drivers of floral trait diversity. However, the relative importance of divergent pollinator-mediated selection versus neutral processes in floral character evolution has rarely been explored. We tested for adaptive floral trait evolution by comparing differentiation at neutral microsatellite loci to differentiation at quantitative floral, growth, and phenological traits in an *Ipomopsis aggregata* hybrid zone. *I. aggregata* subsp. *candida* displays slender white tubular flowers that are typical of flowers pollinated by hawkmoths and subsp. *collina* displays robust red tubular flowers typical of flowers pollinated by hummingbirds; yet hybrid flower morphs are abundant across the East Slope of the Colorado Rockies. We estimated genetic differentiation (F_{st}) for nuclear and chloroplast microsatellite loci and used a half-sib design to calculate trait divergence (Q_{st}) for quantitative traits from collection sites across the morphological hybrid zone. We found very little evidence for population structure, weak isolation-by-distance, and estimated global F_{st} to be 0.032. Q_{st} values for several floral traits including corolla length, corolla width, color, and nectar volume were large and significantly greater than global F_{st}. Additionally, we performed multivariate comparisons of neutral loci to genetic correlations within and between populations and found a strong signal for divergent selection, suggesting that specific combinations of floral display and reward traits may be the target of selection. Our results indicate that there is little support for historical subspecies categories in this hybrid zone, yet floral traits are more diverged than expected due to drift alone. Non-neutral divergence for multivariate quantitative traits suggests that selection by pollinators is maintaining a correlation between display and reward traits.

Introduction

Flowers are extremely diverse and vary in color, scent, size, shape, and pollination reward. These traits are thought to have evolved principally as adaptations for attracting and exploiting pollinators, which in turn evolve characteristics improving their ability to forage on particular flowers. An important aspect of this co-evolutionary process is that it may lead to phenotypic and genetic divergence among plant populations adapted to different pollinator communities and eventually speciation (Coyne & Orr, 2004; Fenster et al., 2004; Grant & Grant, 1965; Grant, 1949, 1981; Stebbins, 1970; Waser et al., 1996) The predominant view is that adaptation to the most abundant or efficient pollinator in an isolated population results in floral divergence, and that subsequent pollinator preference or mechanical isolation will limit intercrossing if diverged populations come into secondary contact (Fenster et al., 2004; Grant, 1949, 1981, 1992a, 1992b; Waser et al., 1996). Co-evolutionary change coupled with the development of pollinator isolation and divergence is arguably one of the most important processes generating diversity in flowering plants (Grant, 1994).

The vast majority of plant adaptation studies focus on single traits and univariate analyses. However, organisms are composed of many interrelated phenotypes that are often integrated and function jointly to impact individual fitness (Kirkpatrick, 2009; Walsh & Blows, 2009). For example, many floral traits are likely to interact with respect to overall size and shape and how they attract pollinators and influence the efficiency of pollinator visitation (Faegri & Pijl, 1979; Fenster et al., 2004; Grant & Grant, 1965). Pollinator effectiveness is likely associated with appropriate matching of corolla tube length, stigma and anther positioning, and the specific characteristics of visiting pollinators. Likewise, features associated with pollination syndromes suggest strong associations between display and reward characters. A classic example is the commonly observed associations between display characteristics, such as flower color, and nectar

rewards for either bird or moth pollination systems. The hummingbird pollination syndrome is characterized by short and robust red corolla tubes, little fragrance, and copious production of dilute nectar. In contrast, the hawkmoth pollination syndrome is characterized by a long and delicate pale corolla tube, strong fragrance, and low nectar production. These patterns suggest that pollinators may impose strong correlational selection (Campbell, 2009; Lande & Arnold, 1983; Sinervo & Svensson, 2002) on functional floral traits, possibly impacting the evolution of floral trait genetic covariances through either the accumulation of pleiotropic effects or the build-up of linkage disequilibrium. Despite these interesting features, few studies of pollinator-mediated selection and floral adaptation have embraced more sophisticated multivariate analyses (Campbell, 2009).

A longstanding goal in plant biology has therefore been to elucidate the factors that constrain or facilitate floral adaptation and that drive pollinator-mediated phenotypic divergence. A number of evolutionary mechanisms can influence phenotypic divergence, and a thorough understanding requires information on the relative importance and consequences of such forces on appropriate spatial and temporal scales (Wade & Goodnight, 1998). For example, genetic drift occurring in small isolated demes will lead to a loss of genetic variability within demes, and neutral divergence among demes (Lande, 1975, 1992; Wright, 1931). Recurrent polygenic mutation will supply new genetic variability to populations and lead to neutral and potentially adaptive divergence among populations (Lande, 1992; Lynch & Hill, 1986). Gene flow, in contrast, will tend to homogenize demes and may limit local adaptation through gene swamping, migration load, or migration meltdown (Lenormand, 2002 and papers reviewed therein). Natural selection may result in either inter-deme divergence or convergence depending on the form of selection. Many field studies of pollination biology have been conducted within a single sampling location and therefore may miss the complexity or dynamics imposed by population structure and multiple evolutionary forces (Hoeksema & Forde, 2008; Thompson, 1994, 1999; Waser et al., 1996). Consequently, studies integrating across spatial scales and evolutionary processes hold promise for providing a richer

understanding of phenotypic change driven by plant-pollinator interactions (Nuismer et al., 1999; Thompson, 1994; Wade & Goodnight, 1998). Notably, Hopkins et al. (2011) and Streisfield and Kohn (2005) use neutral loci to test for, and find evidence of selection on flower color.

Comparison of the divergence of quantitative traits and neutral genetic markers provides one method for integrating across spatial and temporal scales and for testing the relative importance of neutral versus adaptive processes. This approach, known as Q_{st}-F_{st} comparison, is based on evaluating the degree of within to among population genetic differentiation at both putatively neutral markers and hypothesized adaptive quantitative traits. The genetic differentiation at molecular markers is commonly quantified using the metric F_{st}, which at neutral loci is largely influenced by effective population size and the degree of among population migration over evolutionary timescales (Wright, 1951). Q_{st} is an analogous metric of the genetic differentiation within and among populations for a quantitative trait that has an additive genetic basis. Q_{st} for a neutrally evolving quantitative trait is expected to be, on average, equal to F_{st} for a neutral genetic locus (McKay & Latta, 2002; Spitze, 1993; Whitlock, 2008) As such, accurate measures of F_{st} at neutral markers can provide a useful null hypothesis for inferences of evolutionary forces acting on a quantitative trait. For example, if empirical measures of $F_{st} \approx Q_{st}$ we can infer that the observed quantitative trait differentiation could be explained by genetic drift/gene flow balance alone. If Q_{st} is significantly less than F_{st}, indicating less divergence among populations than expected by genetic drift/gene flow balance alone, stabilizing selection may be acting to maintain the phenotype across populations. If Q_{st} is significantly greater than F_{st}, indicating that the focal trait has diverged among populations more than expected by genetic drift/gene flow balance alone, divergent selection may underlie the observed phenotypic divergence. This approach was pioneered by Spitze (1993) and despite the difficulty in accurately measuring F_{st} and Q_{st} (O'hara & Merilä, 2005; Whitlock, 2008) the past decade has seen an exponential increase in refinement of the method and application to empirical studies (Gilbert & Whitlock, 2014; reviewed in: Le Corre & Kremer, 2012; Whitlock & Guillaume, 2009). A logical

extension of this method incorporates the covariance structure of individual components of a complex trait. The genetic architecture of complex traits can shed light on the selection pressures that shape phenotypic divergence. In this study, we are interested in the response to selection by particular components of complex floral phenotypes.

Scarlet gilia (*Ipomopsis aggregata* Polemoniaceae) is an ecologically well-studied wildflower with a wide geographical range across most of the western United States. The *I. aggregata* species complex is composed of at least eight subspecies, several occurring sympatrically, with clear contact zones and regions of hybridization (Grant, 1956; J. M. Porter & Johnson, 2000; Wolf & Soltis, 1992). Co-evolutionary interaction with pollinators (hummingbirds, hawkmoths, bees, butterflies) is likely an important driver of diversification and hybridization within the *I. aggregata* complex (Grant, 1992a). Field studies of *I. aggregata* subsp. *aggregata* arguably provide the best evidence for the occurrence and strength of pollinator-mediated selection on floral morphology in natural populations. For example, long-term studies at the Rocky Mountain Biological Laboratory (RMBL) in western Colorado, have documented that pollen-limitation coupled with variation in pollinator visitation and efficiency can impose strong natural selection on floral characters [flower size and shape (Campbell, 1989, 1991, 1996), inflorescence display (Campbell, 1991; Mitchell, 1994), nectar production (Mitchell & Waser, 1992; Mitchell, 1993), and timing of male and female reproduction (Campbell et al., 1996; Campbell, 1991, 1997)]. Studies of an *Ipomopsis* hybrid zone at RMBL (*I. aggregata ssp. aggregata* x *I. tenuituba*) have also detailed variable selection imposed by hummingbird and hawkmoth pollinators on floral morphology (Campbell et al., 1997, 1998; Campbell, 2003; Melendez-Ackerman & Campbell, 1998; Melendez-Ackerman, 1997). Surprisingly little work on plant-pollinator interaction has been pursued in other *I. aggregata* populations, subspecies, or hybrid zones.

We extend earlier studies by exploring patterns of molecular marker and quantitative trait differentiation occurring on the East Slope of the Rocky Mountains. In particular, we evaluate the population structure and floral trait differentiation in local occurrences of *I. aggregata* subsp. *candida* (Rydb.) V.E. Grant & A.D. Grant and *I.*

aggregata subsp. *collina* (Greene) Wilken & Allard (hereafter referred to as *candida* and *collina*) using univariate (Q_{st}-F_{st}) and multivariate (genetic covariance matrix deconstruction) comparisons to test for departure from expectations of neutral evolution. The flowers of subspecies *candida* exhibit classic hawkmoth pollination syndrome characteristics. They are delicate, with relatively long and narrow corollas, inserted anthers, and range in corolla color from pure cream white to light shades of pink (Figure 1.1a). *I. a. candida* is primarily pollinated by white-lined hawkmoths, *Hyles lineata* (Sphingidae) and, to a lesser degree, by resident and migratory hummingbirds (Elam & Linhart, 1988). In contrast, the flowers of subspecies *collina* exhibit classic hummingbird pollination syndrome characteristics. They are robust, with relatively wide and short corollas, exerted stigmas, red flowers, and are primarily pollinated by hummingbirds (Figure 1.1b). Nectar characteristics are important floral rewards but are rarely studied. *Collina* flowers typically produce a large amount of dilute nectar in comparison to *candida* flowers, again matching the typical hummingbird and hawkmoth associated suites of traits (Fenster et al., 2004; Mitchell, 2004).

I. a. candida and *collina* hybridize across much of the East Slope of the Rockies (Figure 1.2). This hybrid zone was first investigated with respect to the introgression of floral characters by Wilken and Allard (1986), and Grant and Wilken (1987) and Elam and Linhart (1988). One hypothesis is that the *candida* morph diverged from a *collina*-like ancestor during the last glacial maximum as an isolated allopatric population at low elevation on the shortgrass prairie. Following glacial retreat and climate change, *candida* populations are thought to have extended their range from the prairie to higher elevations and thus come into secondary contact with *collina* (Grant & Wilken, 1987), generating a complex mosaic of hybridization and introgression throughout the East Slope. Given this hypothesis, we would predict both neutral and quantitative trait divergence, depending on the degree and extent of isolation and reproductive incompatibilities. Alternatively, *candida* and *collina* may be undergoing primary divergence ostensibly by pollinator-mediated disruptive selection on floral types (Wolf & Soltis, 1992). This scenario may

result in substantially stronger quantitative trait divergence with potentially little or no neutral marker divergence because of ongoing gene flow.

In this study we use a two-pronged approach to address differentiation across the broad geographical distribution of *I. aggregata* on the Front Range of the Rockies. First, we utilized putatively neutral microsatellite markers to characterize molecular genetic differentiation, population structure, and isolation-by-distance across occurrences in the Front Range. Second, we used a greenhouse common garden study to estimate quantitative genetic structure and evaluate population differentiation of floral, phenological, and growth traits. Together, these data allowed inference on the potential adaptive differentiation of floral characters across a broad geographical region and shed light on earlier hypotheses concerning the *candida/collina* hybrid zone.

Methods

Study System and Sampling Locations

Scarlet gilia is a common monocarpic, perennial wildflower of western montane habitats. It is frequently found in large grassy meadows and along roadsides interspersed with small patches of ponderosa pine forest occurring from 1740 m to over 3050 m elevation on the East Slope of the Rocky Mountains in Colorado (Wilken & Allard, 1986). After germination, scarlet gilia grows as a hardy rosette for 2–5 years until it bolts an indeterminate flowering stalk, reproduces, and dies. Scarlet gilia has showy inflorescences with flowers that are hermaphroditic, protandrous, and self-incompatible. The flowers of subspecies *candida* and *collina* and their hybrids range in color from pure creamy white, light shades of pink, to vivid scarlet red (Figures 1.1a-b, 1.2). They are pollinated by the white-lined hawkmoth (Sphingidae: *Hyles lineata*) and, to a lesser extent, by resident and migratory hummingbirds, including Broad-tailed Hummingbirds (*Selaphorus platycerus*), Rufous Hummingbirds (*Selaphorus rufus*), and Calliope

Hummingbirds (*Stellulla calliope*) (Elam & Linhart, 1988) (T. Juenger, A. Kenney & E. Milano, personal observation).

We studied 39 occurrences of scarlet gilia from locations in Colorado (Table A1, Figure A1). The occurrences were identified from prior studies (Grant & Wilken, 1987; Wilken & Allard, 1986) along with new locations that we identified. Sampled locations covered low elevation foothills, montane grasslands, and mixed forest habitats in the Northern and Southern Front Range, Sangre De Cristo Range, and Wet Mountains.

Neutral Genetic Markers

At 22 locations (Figure 1.3, Table A1), flowers and leaf tissue were collected from 8-12 individual plants. Genomic DNA was extracted from leaf tissue using a DNeasy Plant tissue kit (Qiagen). Genomic DNA was genotyped at 11 nuclear microsatellite loci derived from 4 custom enrichment microsatellite libraries (Stearns et al., 2008). This set of microsatellites covers 6 of 7 *I. aggregata* chromosomes, as determined by standard linkage analysis (Campitelli et al., in prep). In addition, each sample was genotyped at 5 chloroplast microsatellites (Provan et al., 2001). Polymerase chain reaction (PCR) amplification was carried out in a Peltier DNA Engine Tetrad 2 thermal cycler (BioRad) in 5uL of reaction mixture containing 20 ng of template DNA, ddH_2O, 0.8 μM dNTPs, 1 × NH_4 reaction buffer (Bioline), 1.5 μM MgCl2, 0.01 μM forward primer with a 19 bp M13F tail (CACGACGTTGTAAAACGAC), 0.15 μM reverse primer, 0.15 μM M13- HEX, FAM or Tamara, and 0.01 U of *Taq* DNA Polymerase (Bioline). DNA was initially denatured at 94.0 °C for 2 min, cycled 35 times at 94.0 °C for 30 s, 55 °C for 30 s, and 72 °C for 45 s, and then finished with a final extension step at 72 °C for 10 min. Amplification products were detected on an Applied Biosystems 3100 DNA Sequencer (Life Technologies), and fragments called and analyzed using GeneMarker (v. 1.3) software.

Quantitative Phenotypic Traits

A dam half-sib design was implemented for our quantitative genetic studies by field collecting open-pollinated fruits from 182 maternal plants from 17 locations (Figure A1, Table A1). These seeds were cold stratified for 6 weeks in shallow trays of sand in a cold room (4 °C) at the University of Texas at Austin. Seeds were then placed in the UT Austin greenhouse for germination. Once true leaves were present, seedlings were transplanted into cone-tainers (Steuwe & Sons, SC10, 164ml volume) with a 3:3:4 ratio of sand, turface (Profile Products LLP), and Pro-mix (Premier Tech) as a potting medium and randomized across growing trays and benches. Plants were grown under long day conditions (16h day/8h night) using supplemental overhead high pressure sodium lighting and standard greenhouse conditions based on evaporative cooling. After 4 months of vegetative rosette growth, plants were returned to a 4° C cold room for 4 months of vernalization to simulate a natural overwintering. Following cold vernalization, plants were subsequently transferred to the greenhouse for their final growing season. Plants bolted a growing inflorescence and flowered over a 158-day period (average 67.7 +/- 0.96 SE days) following reintroduction to the greenhouse. In total, 745 plants were established for our greenhouse studies forming 180 maternal sibships across 17 collection locations.

Siblings were phenotyped for a series of traits associated with floral display, reward, and phenology. Display traits and reward traits were chosen to best characterize the variation in *candida/collina* flowers. Display traits measurements included corolla tube length, corolla tube width, longest petal length, and flower color on three flowers per plant. The color variation observed in *Ipomopsis* flowers is primarily the product of anthocyanin pigments (Harborne & Smith, 1978). Pigments from harvested corolla tubes were extracted with methanol and quantified on a standard plate reader (Beckman DTX880) at a wavelength of 492, the major absorbance wavelength of the primary anthocyanin pigment, pelargonidin. Dry weight of the floral tissue was used to standardize the absorbance intensity as a quantitative estimate of color. Reward trait measurements included nectar volume and concentration on three flowers per plant. To

standardize measurements and avoid evaporation, elongating buds were taped shut in the afternoon the day before they were to open. Forty-eight hours later, each flower was carefully removed and nectar was extracted from the base of the corolla with a capillary tube (Drummon Scientific). Nectar concentration was quantified using a hand-held refractometer (Eclipse 45-03, Bellingham and Stanley Ltd.). All experimental plants were well-watered to reduce the potential impact of water limitation on nectar production.

Data was also collected for two traits associated with general growth and phenology. Flowering time was measured as the number of days of growth in the greenhouse following vernalization, to the date of the first open flower. *Ipomopsis* grows as a rosette that develops a hardy rootstock before flowering. Rootstock diameter is a strong predictor of overall plant size and is correlated with plant height and flower production (Juenger & Bergelson, 1998). Rootstock diameter was measured at the end of the experiment after 7 months of growth in the greenhouse.

Population Genetic Data Analysis

Descriptive statistics for nuclear and chloroplast microsatellites were generated using the R package *adegenet* (Jombart, 2008, version 1.3-4) based on allele designations from fragment size determinations. Here we provide a brief summary of the microsatellite data used for population genetic analysis. Across the entire sample of individuals (n=230), the mean number of alleles at nuclear microsatellites was 46 +/- 6.72 SE; the mean observed heterozygosity was 0.633 +/-0.042 SE; and the mean inbreeding coefficient (F_{IS}) was 0.257 +/- 0.034 SE. After Bonferroni correction only two populations had one locus each that exhibited a departure from Hardy-Weinberg expectations. Population structure was assessed by principal component analysis (PCA) and discriminant function analysis (DAPC) in *adegenet* (Jombart & Devillard, 2010) and with the Bayesian clustering program STRUCTURE (Pritchard et al., 2000). STRUCTURE analyses were completed with a model specifying admixture, no location prior, a 10,000 iteration burn-in, and a 20,000 iteration MCMC. Results were evaluated

with Structure Harvester and optimal K determined using the delta method (Earl & vonHoldt, 2012; Evanno et al., 2005). Isolation-by-distance was evaluated using a Mantel test to assess the correlation between a matrix of Nei's pairwise F_{st} and Euclidian geographic distance.

Quantitative Genetic Analysis

The quantitative phenotypic traits were box-cox transformed to meet the requirement of normally for all analyses. Traits were also assumed to have purely additive modes of inheritance and we did not make any assumptions about historical sub-species divisions. Q_{st}-F_{st} comparisons must be completed using a rigorous statistical test (Merilä & Crnokrak, 2001; O'hara & Merilä, 2005); in particular, a key assumption is that the distribution of neutral Q_{st} markers should be equal to that of neutral F_{st} markers (Whitlock, 2008). Whitlock and Guillaume (2009) developed a method using parametric simulations and bootstrapping to predict the null distribution of neutral Q_{st} and compare the observed Q_{st} to the tails of the simulated null distribution. An extension of this procedure allows for an unbalanced open-pollinator design to calculate observed Q_{st} (Gilbert & Whitlock, 2014). This method allows for a maternal pedigree of half-sib design with no need to know sire, which is a common design for outcrossing plant populations. We centered our analyses on a half-sib design based on a study that found multiple paternity in over two-thirds of *Ipomopsis* fruits, with number of sires averaging 4 per fruit (Campbell, 1997). Calculation of F_{st}, Q_{st}, and associated distributions and confidence intervals were completed using the R package *QstFstComp* provided by Gilbert and Whitlock (2014).

Multivariate analysis

We first performed a PCA on phenotypic traits to visualize differences between collection sites. We then performed two multivariate methods to test for selection. The

first is a natural extension of the Qst-Fst comparison using genetic covariance matrices within and between populations to test for non-proportionality (Martin et al., 2008). The second uses existing neutral markers and quantitative traits to simultaneously estimate ancestral population parameters and derive theoretical null expectations for each population against which to test for signatures of divergent or stabilizing selection in observed populations (Ovaskainen et al., 2011).

We performed a multivariate extension of the Q_{st}-F_{st} analysis by comparing the among population genetic covariance matrix ($\mathbf{G_B}$) to the average within population genetic covariance matrix ($\mathbf{G_W}$) using the two-part test developed by Martin et al. (2008). For the multivariate analyses we use the term 'population' to refer to the collection sites. We used a MANOVA to estimate $\mathbf{G_W}$ and $\mathbf{G_B}$ from genetic family means. Under neutral expectations $\mathbf{G_W}$ and $\mathbf{G_B}$ are expected to be proportional where the proportionally coefficient, $\rho_{st} = 2*F_{st} /(1-F_{st})$ (Martin et al., 2008; Rogers & Harpending, 1983). In this two-part test we calculated a proportionality coefficient (ρ_{st}) such that $\mathbf{G_B} = \rho_{st} * \mathbf{G_W}$. We first tested if ρ_{st} for quantitative traits ($G\rho_{st}$) was significantly greater than ρ_{st} calculated from the neutral marker loci ($N\rho_{st}$). If $G\rho_{st} > N\rho_{st}$ then we inferred deviation from neutral expectations. We then tested ρ_{st} itself using a likelihood ratio test where H_o: $\mathbf{G_B} = \rho_{st} * \mathbf{G_W}$, and H_1: $\mathbf{G_B}$ and $\mathbf{G_W}$ are completely unrelated. We used a Bartlett correction of the likelihood ratio to account for small degrees of freedom. This second part of the test gives insight into the relative structure and orientation of $\mathbf{G_B}$ and $\mathbf{G_W}$. Calculations were performed using customized R scripts from Martin et al. (2008) and we applied a matrix bending correction to $\mathbf{G_B}$ to make it positive-definite. Finally, we performed eigen deconstruction to compare trait loadings on the major principal components of $\mathbf{G_B}$ and $\mathbf{G_W}$.

The second multivariate test to detect signatures of divergent or stabilizing selection in quantitative traits is similar to the animal model in that we assumed that the additive variance among relatives, as measured by the animal model, holds at the population level (Ovaskainen et al., 2011). The genetic value of an individual now becomes the genetic mean of a population and the coancestry coefficient between

individuals is now the coancestry coefficient between individuals belonging to different populations. This test, developed by Ovaskainen et al. (2011), is based on the assumption that all populations are derived from a common ancestral population. We use Bayesian inference to estimate a genetic coancestry matrix based on neutral markers and a genetic covariance matrix based on quantitative traits. We then use these ancestral matrices to derive expectations of evolution under a null model of random genetic drift for each population. This is in contrast to the previous test where we averaged F_{st} and $\mathbf{G_W}$ across populations. We then explicitly test for neutrality using the likelihood of the observed pattern given the theoretical expectation based on random genetic drift. The signature of selection test statistic, S, is defined as the probability that the actual pattern of additive genetic divergence is less than that of the randomized realizations of additive genetic divergence. A value of S≈1 means that the observed pattern is unlikely under neutrality. The typical pattern under neutrality is S≈0.05. In the case of stabilizing selection, the value of actualized divergence is lower than expected due to random drift so S≈0. To perform this test we used a subset of 11 populations that had both neutral marker and quantitative trait information. (Table A1, Figure A1). The analyses were completed using the R packages *RAFM* to calculate the posterior distributions of drift and gene flow from each population, and *driftsel* to calculate the expected differences in population trait means and perform the neutrality test using 10,000 random draws (Karhunen et al., 2013).

Results

Population genetic structure

Neutral marker analyses revealed little genetic differentiation or structure across the region. Overall, pairwise nuclear F_{st} is low (average pairwise F_{st} = 0.051 +/- 0.010 SE) but increases with geographic distance, indicating isolation-by-distance (IBD)

primarily along a north-south axis (Figure A2). A Mantel test for IBD reveals a significant correlation between genetic and geographic distance matrices for nuclear markers ($R^2 = 0.623$; $p < 0.001$; 9999 permutations) and a weaker yet significant pattern for chloroplast markers ($R^2 = 0.290$; $p = 0.0065$; 9999 permutations, Figure A3). Maternally inherited chloroplast microsatellite markers exhibited relatively more genetic differentiation across the sampled locations. Average cp-F_{st} (haplotype diversity) between pairs of populations was 0.451 +/- 0.215 SE. Results from a Discriminant Analysis of Principal Components (*adegenet*) found K=2 clusters that do not correspond to putative sub-species divisions, instead suggesting a geographic trend from north to south (Figure 1.3). STRUCTURE did not find a significant number of K>1 clusters. A PCA of the molecular data showed little differentiation between the populations (Figure 1.4a, Figure A4), whereas a PCA of floral traits revealed distinct clusters of populations (Figure 1.4b). These results suggest a high degree of nuclear gene flow between locations and are consistent with relatively long range pollen dispersal by animal pollinators, large effective population sizes, and little reproductive isolation among floral syndromes.

Q_{st} and F_{st} comparison

We found Q_{st} for several floral traits to be greater than both average pairwise F_{st} (0.051 +/- 0.010 SE) and global F_{st} (0.032, 95% CI: 0.023-0.042), indicating that the traits are significantly more differentiated between populations than expected from drift alone (Figure 1.5). The floral display and reward traits with significant Qst values are corolla width, corolla length, flower color, and nectar volume. We did not find significant among population genetic variation for petal length or nectar concentration. Additionally the phenology traits, root diameter and flowering time show Q_{st} signals consistent with neutral evolution. The conclusions drawn from our Q_{st}-F_{st} comparisons are robust to analysis assumptions based on sibship relatedness. Q_{st} values for corolla length, corolla width, corolla color, and nectar volume were consistently significantly larger than F_{st} whether our dam design analysis considered families to be composed of half (relatedness

= 0.25) or full (relatedness = 0.50) sibships. Our conclusions are also robust to an unbalanced sampling design. Our original Qst dataset had an excess of populations without matching neutral marker information and likewise for the F_{st} dataset (Figure A1). We removed the unmatched populations from the analysis (indicated in Table A1) and found the conclusions to be qualitatively consistent with those of the full dataset (data not shown).

Multivariate Tests for Neutrality

We extended the univariate Q_{st}-F_{st} test into multivariate trait space by incorporating genetic variance-covariance architecture of the floral traits that are otherwise lost in a univariate analysis. We estimated the among ($\mathbf{G_B}$) and average-within ($\mathbf{G_W}$) population covariance matrices and found stark differences in the structure of genetic covariation (Table 1.1). The dominant covariance structure among populations ($\mathbf{G_B}$) was indicative of the expected pollination morphotypes; either red, short, and wide flowers with abundant nectar or white, long, and thin flowers with sparse nectar. In contrast, the major components of quantitative trait variation within populations ($\mathbf{G_W}$) were petal length and nectar concentration. We confirmed these results with a two-part test for departure from neutral expectations using a proportionality test statistic, ρ_{st}. We first tested if ρ_{st} for quantitative traits ($G\rho_{st}$) was significantly greater than ρ_{st} calculated from the neutral marker loci ($N\rho_{st}$) and then we tested if ρ_{st} itself was a valid coefficient of proportionality or if $\mathbf{G_B}$ and $\mathbf{G_W}$ were completely unrelated. We found $G\rho_{st}$ (0.43, 95% CI: 0.33-0.61) was significantly greater than $N\rho_{st}$ (0.066, 95% CI: 0.050-0.084) so we rejected proportionality between $\mathbf{G_B}$ and $\mathbf{G_W}$ (Bartlett corrected p-value<0.0001). An eigen-analysis of the two genetic matrices illustrated the non-proportionality of the covariance structure (Figure 1.6). We observed large differences in both the amount of variation described by the first two principal component (PC) axes, and the trait loadings on each. PC 1 and PC 2 described almost 100% of the variation in $\mathbf{G_B}$ and were comprised of the same 4 significant Q_{st} traits. However, PC 1 and PC 2 described only

61.5% of the variation in $\mathbf{G_W}$, and were primarily comprised of petal length, nectar concentration and to a lesser extent, nectar volume. The correlation coefficient (r) between $\mathbf{G_B}$ and $\mathbf{G_W}$ PC1 is -0.92 and for PC2 is -0.33.

In our second multivariate approach we estimated a population level coancestry matrix (Figure 1.7a), based on neutral loci, and an ancestral additive genetic covariance matrix based on neutral expectations of quantitative traits. We then tested each observed population against a null model based on theoretical expectation under random genetic drift and calculated a selection statistic, S where $S < 0.05$ indicates stabilizing selection and $S > 0.95$ indicates directional selection. We found strong evidence for divergent selection across populations ($S = 1.0$, Ovaskainen et al. 2011) for a multivariate model that included all 6 floral traits. We visually represent observed trait divergence between populations in relation to the estimated ancestral population (A), and expectation under drift for three pairs of traits. Populations that fall inside their respective 50% confidence ellipses are exhibiting neutral divergence or possible stabilizing selection whereas populations outside of their corresponding ellipses are exhibiting directional selection. We found that the majority of populations exceeded expectations of neutral divergence for trait pairs including corolla length, corolla width, color, and nectar volume (Figures 1.7b-c) indicating evidence for multivariate divergence. However, most populations fell inside the 50% neutrality ellipse for petal length and nectar concentration (Figure 1.7d).

Discussion

Plant-pollinator coevolution is thought to be a primary driver of plant population divergence and ultimately plant speciation. Here, we explored the population structure and patterns of quantitative genetic divergence in floral traits in a putative *I. aggregata* hybrid zone. We found little genetic differentiation or population structure along traditional subspecies designations. Overall, global nuclear F_{st} was low and genetic PCA, discriminant function, and Bayesian clustering approaches found no clear groupings that

are strongly coincident with corolla color or other floral morphological traits. At best, nuclear microsatellite markers exhibited isolation-by-distance that is primarily associated with a latitudinal axis. Together, these data suggest *I. aggregata* demes in the Front Range of the Rockies experience high levels of gene flow and little reproductive isolation among demes, including between demes with considerably diverged suites of floral traits. This molecular data also suggests that the dominant pollinators, including hummingbirds and hawkmoths, have low fidelity to color and do not exhibit strong ethological isolation but are rather flexible in their preferences. This is consistent with earlier findings of Elam and Linhart (1988) who observed hummingbirds and hawkmoths visiting all color morphs along the Front Range, and personal observations from our own field work. As such, the data support at best, a very recent and only shallow divergence of ancestral *candida*/*collina* populations prior to widespread admixture or alternatively, a model of primary divergence and polymorphism in sympatry. Either model suggests a rather rapid evolution of the hawkmoth syndrome morph from the presumed hummingbird syndrome ancestor.

One possible mechanism for the rapid evolution of the hawkmoth pollination morph is based on corolla color. The anthocyanin pathway that produces the red color in *Ipomopsis* is a well characterized pathway (Forkmann, 1991; Grotewold, 2006). Evolutionary shifts in flower color are also well documented in many systems (Faegri & Pijl, 1979; Fenster et al., 2004). The anthocyanin pathway has three major branches and color transition rates are asymmetrical. Loss-of-function mutations often lead to branch inactivation and the pathway diverts to another branch or does not make any anthocyanin pigment (Rausher, 2008). If this is the case, it is feasible to experience a rapid shift from producing red flowers to producing white flowers. This however, does not account for the multivariate component of the suite of floral morphological traits associated with different pollinators.

In contrast to our findings in population structure, Q_{st} analyses reveal substantial among-population differentiation in floral traits including corolla tube length, width, color, and nectar volume. Q_{st} values estimated from common garden studies were

significantly larger than both nuclear and chloroplast F_{st} values, supporting a model of adaptive among-population divergence. Multivariate analyses further support the notion of adaptive differentiation of floral traits. In particular, comparisons of the average within ($\mathbf{G_W}$) and among population ($\mathbf{G_B}$) genetic covariance matrices rejected proportionality – an expectation of neutral divergence (Martin et al., 2008). Similarly, analyses that used inferred ancestral population states to test for divergence beyond that expected from neutral processes provide strong evidence for selective divergence in multivariate trait space. Collectively, these results suggest that suites of floral traits have been the targets of selection – a hypothesis that would be supported by the observation of correlational selection on floral traits in the field.

Response to selection is not just controlled by genetic variation but by genetic covariation (Blows & Hoffmann, 2005; Lande & Arnold, 1983), and defined by the principal components of the genetic covariance matrix (Blows, 2007). In this study we found non-proportionality for $\mathbf{G_W}$ and $\mathbf{G_B}$, implying not only that divergence between populations is not occurring at the neutral rate, but also is different across spatial scales. The first two PCs in $\mathbf{G_B}$ accurately describe the phenotypic divergence we observe in floral traits between *candida* and *collina*. However, the first two PCs for the average $\mathbf{G_W}$ matrix describe much less of the variation than $\mathbf{G_B}$ and are comprised primarily of traits that have little genetic variation. This implies response to selection at the within deme spatial scale is different than response to selection across the range. The results from the molecular markers indicate that this is one large admixed population so we should expect the overall *candida*/*collina* hybrid zone to respond in the dimensions of $\mathbf{G_B}$. This discrepancy between genetic covariance matrices suggests population structure and admixture estimates are important considerations in designing experiments that properly sample the diversity of a focal gene pool. If we had only studied a single hybrid deme, our interpretation of response to selection in the *candida*/*collina* hybrid zone might have been substantially different.

In this case, hummingbird pollinator-mediated selection on color or other display traits may be contingent on covariances with other related floral traits like corolla tube

length or width. Similarly, one expectation in the coevolution of pollination syndromes is the evolution of co-inheritance of display and reward in the partner plant species – display traits must faithfully signal the rewards to pollinators in order to maintain strong interactions. We found that the major genetic axis of the among-population genetic covariance matrix corresponded to strong genetic covariances between corolla tube length and width, corolla color, and nectar abundance. This indicates strong genetic correlations between the display and reward components of the floral phenotype are maintained in the hybrid zone. Genetic correlations can be the result of pleiotropy, physical genetic linkage, or linkage disequilibrium (LD) resulting from correlated selection pressures (Sinervo & Svensson, 2002). Pleiotropy occurs when individual loci have multiple phenotypic effects whereas LD builds up over time due to the non-random association of alleles at different loci. This non-random association can be broken after sufficient random mating and recombination whereas pleiotropic interactions persist. If the genetic correlations between display and reward traits are due to LD then we can assume that they have built up due to pollinator-mediated correlational selection, indicating that pollinator display is an honest indicator of reward. This may support strong adaptive divergence along the expected hummingbird-hawkmoth specialization axes.

Tables

a. $\mathbf{G_W}$

	Corolla Width	Corolla Length	Petal Length	Color	Nectar Volume	Nectar Concentration
Corolla Width	0.2395	-0.0721	-0.0154	-0.0029	0.0173	0.0018
Corolla Length	-0.0721	0.2500	0.0344	-0.0017	-0.0094	-0.0150
Petal Length	-0.0154	0.0344	0.5915	0.0445	0.1163	-0.1399
Color	-0.0029	-0.0017	0.0445	0.1063	0.0409	-0.0097
Nectar Volume	0.0173	-0.0094	0.1163	0.0409	0.3067	-0.1737
Nectar Concentration	0.0018	-0.0150	-0.1399	-0.0097	-0.1737	0.5987

b. $\mathbf{G_B}$

	Corolla Width	Corolla Length	Petal Length	Color	Nectar Volume	Nectar Concentration
Corolla Width	0.6987	-0.5788	0.2772	0.5961	0.6397	0.0533
Corolla Length	-0.5788	0.5761	-0.2004	-0.6112	-0.5518	-0.0606
Petal Length	0.2772	-0.2004	0.1533	0.1169	0.2381	0.0255
Color	0.5961	-0.6112	0.1169	0.9030	0.6060	0.0692
Nectar Volume	0.6397	-0.5518	0.2381	0.6060	0.5982	0.0598
Nectar Concentration	0.0533	-0.0606	0.0255	0.0692	0.0598	-0.0125

Table 1.1 Genetic covariance matrices

Genetic covariance matrices (a) within and (b) between collection sites. Additive genetic variance for each trait occurs along the diagonal, and genetic covariance is mirrored on the off-diagonal. We report the initial genetic covariance matrix estimated between populations ($\mathbf{G_B}$), before positive-definite matrix correction.

Figures

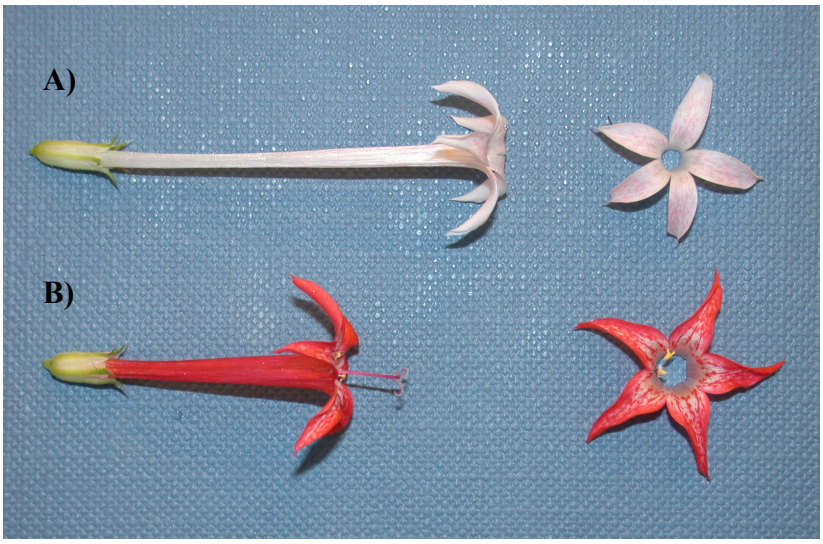

Figure 1.1 *I. aggregata* varieties

Typical (a) *I. a. candida* flowers display long slender white corollas whereas (b) *I. a. collina* flowers display shorter more robust red corollas.

Figure 1.2 *I. aggregata* hybrids

Hybrid *I. a. candida* and *I. a. collina* flowers display a spectrum of color and shape combinations as shown in the field (above) and close-up (below).

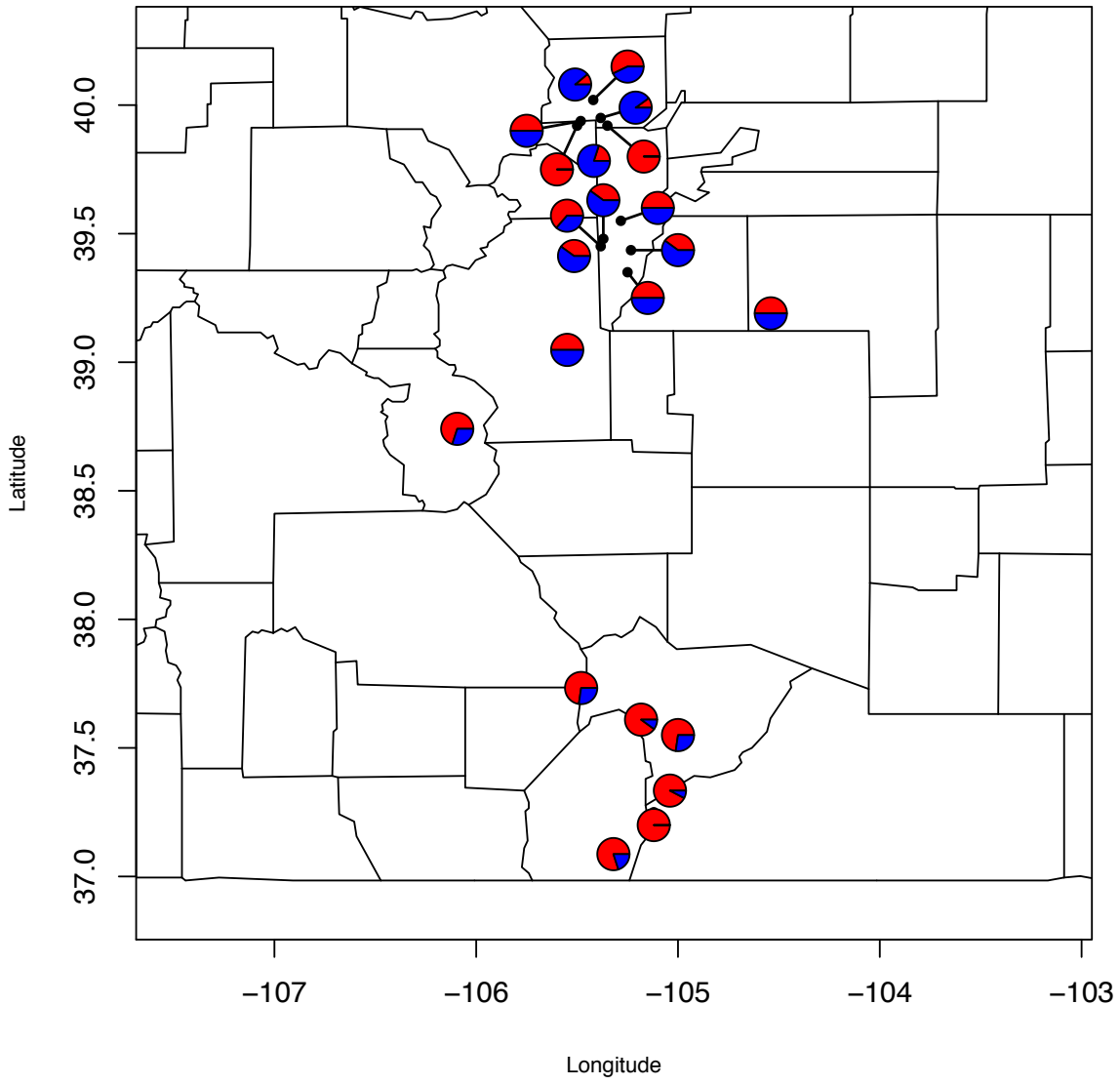

Figure 1.3 DAPC K-means population assignments

Map of collection sites for neutral markers in the Front Range of the Colorado Rocky Mountains with pie charts indicating cluster assignments from DAPC (K=2).

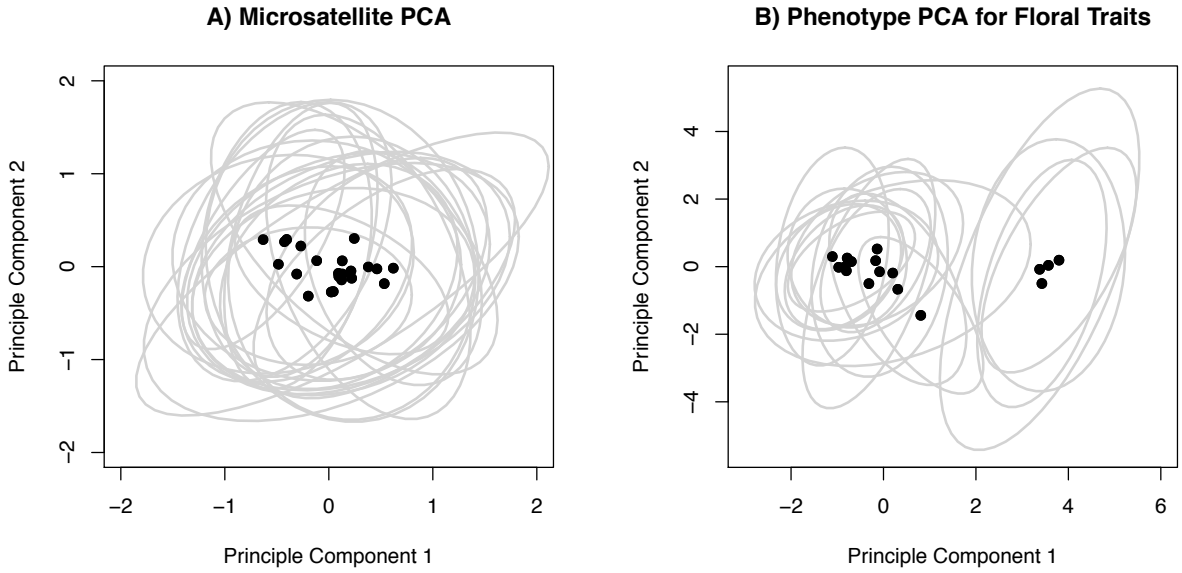

Figure 1.4 PCA for microsatellites and floral traits

Principal components (PCs) 1 and 2 for (a) neutral microsatellite loci and (b) quantitative floral phenotypes. Dots represent population means, ellipses indicate 95% CI.

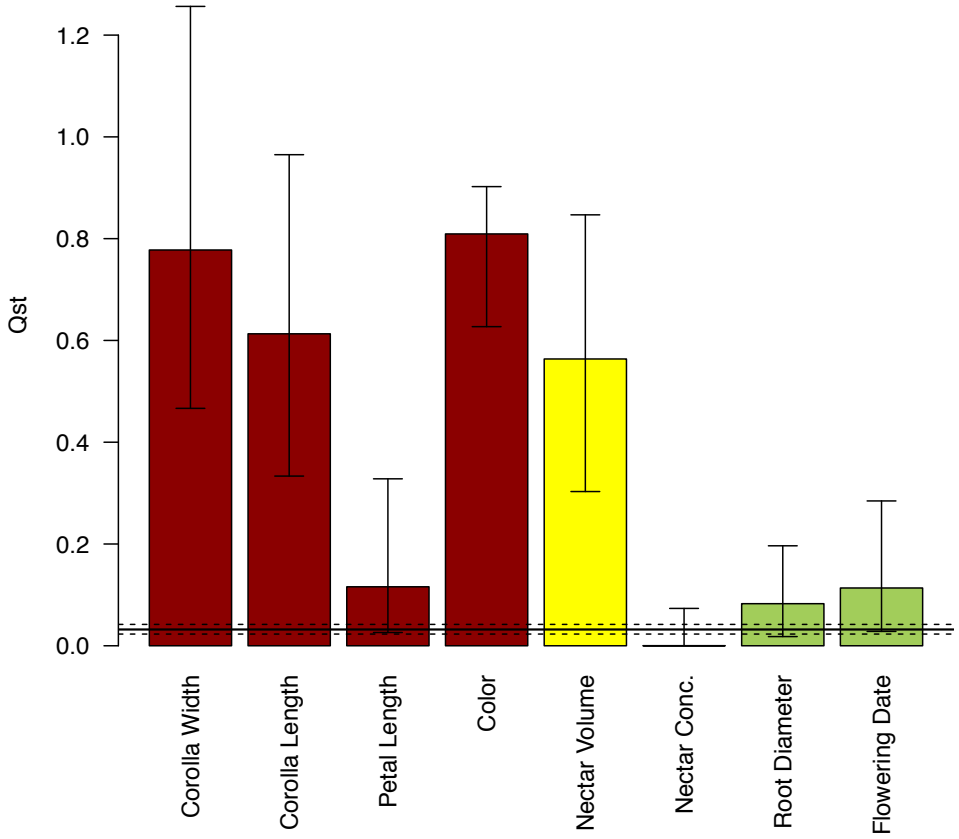

Figure 1.5 Q_{st} estimates for floral traits

Q_{st} and 95% confidence estimates for display (red), reward (yellow), and phenology (green) traits. Global F_{st} (solid) and 95% confidence estimate (dashed) for microsatellite loci indicated by horizontal lines.

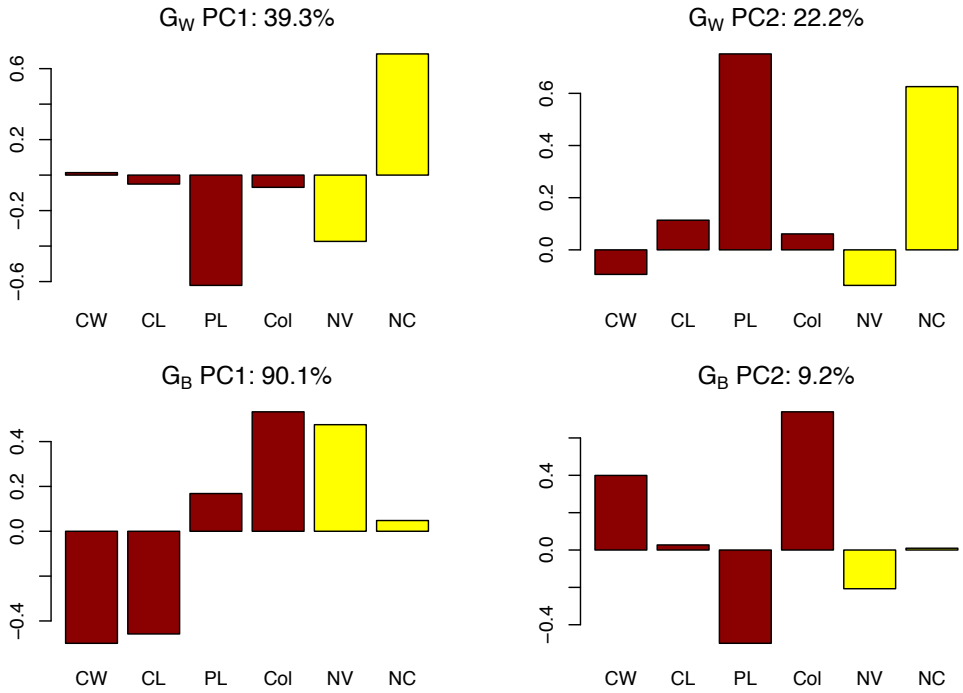

Figure 1.6 G matrix deconstruction

Trait loadings, and percent variance explained by the first two principal components (PCs) for the within ($\mathbf{G_W}$) and between ($\mathbf{G_B}$) genetic covariance matrices. CW= corolla width, CL= corolla length, PL= petal length, Col= color, NV= nectar volume, and NC=nectar concentration.

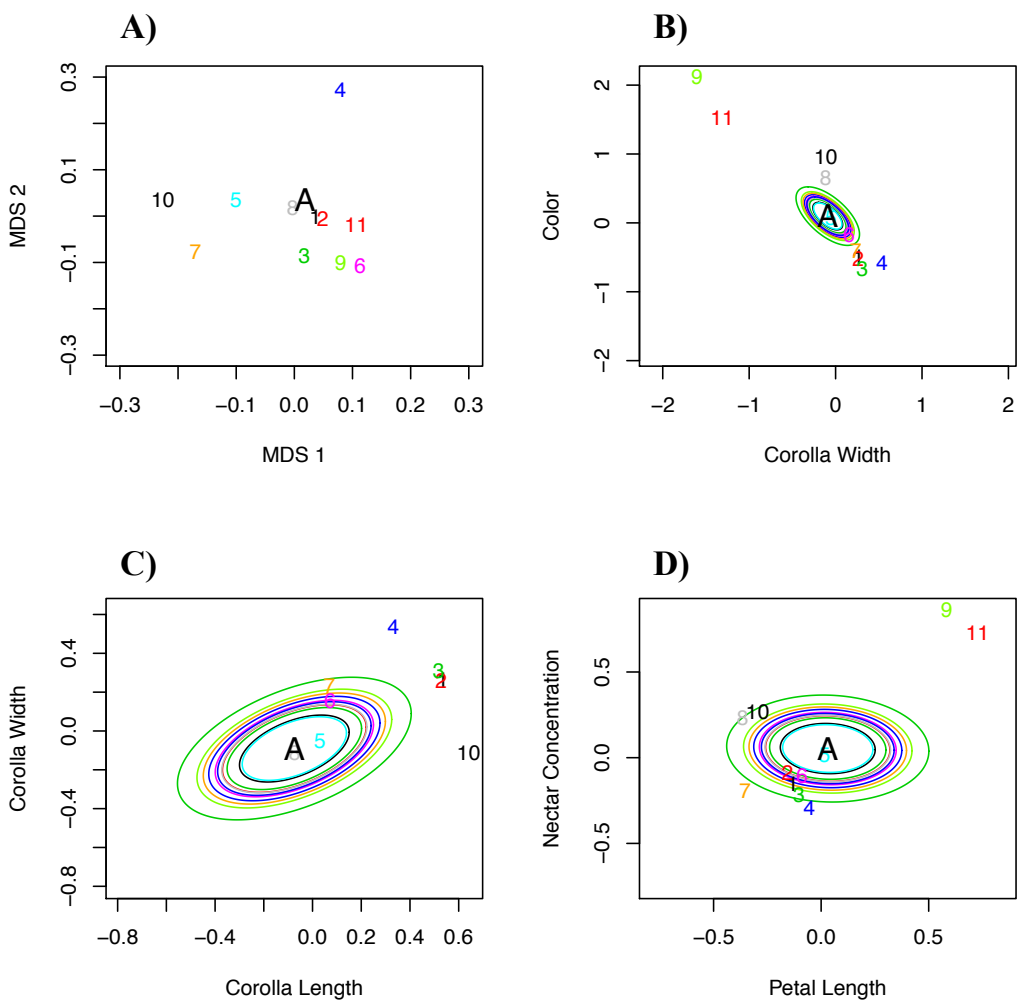

Figure 1.7 *driftsel* analysis

Results from *driftsel* analysis: (a) Neutral genetic distance for each population mean from the inferred ancestor, **A**. Distance calculated as units of ancestral standard deviation and plotted in a 2D plane using multidimensional scaling. Figure produced using modified code from the viz.theta function in *driftsel*. (b-d) Additive genetic population means for selected trait pairs, indicated by population ID. Corresponding color ellipses, centered on inferred ancestral mean, represent the median estimated drift distance for each population. Figure produced using modified code from the viz.traits function in driftsel. Unique population IDs correspond to locations in Table A1.

Chapter 2: The genetic basis of upland/lowland ecotype divergence in Switchgrass (*Panicum virgatum)*

Abstract

Locally adapted ecotypes are a common phenomenon generating plant diversity within species. Despite their widespread occurrence, we know surprisingly little about the genetic mechanisms that lead to locally adapted traits or ecotype formation. The genetic architecture underlying locally adapted traits dictates how an organism will respond to environmental selection pressures with applications in evolutionary ecology, conservation, and crop development. *Panicum virgatum* (switchgrass) is a native perennial grass with two major ecotypes, upland and lowland, which are locally adapted to northern xeric and southern mesic habitats, respectively. We constructed a genetic mapping population segregating ecotypic variation in switchgrass, based on a four-way outbred cross design between two upland and two lowland ecotype accessions, and developed a joint linkage map using ddRADseq markers. QTL were mapped for 9 traits that are generally diverged between ecotypes, including flowering time, plant size, physiological processes, and disease resistance. We found that most of the QTL had intermediate to small effects and clustered on a limited number of linkage groups. We characterized the functional allelic effects of each QTL and found that 60% of the functional allelic effects displayed patterns associated with fixed differences between ecotypes. These results suggest there is considerable standing genetic variation within local populations as well as between ecotypes for a number of ecologically important traits. One third of the QTL had allelic effects inconsistent with the direction of ecotypic divergence, suggesting that potentially maladaptive variation segregates within switchgrass populations. We also found a number of traits that maintained significant correlations after recombination, some of which shared QTL, suggesting possible

pleiotropic relationships. Resolving the genetic architecture of locally adapted traits is not only important for evolutionary ecology but has functional applications in marker assisted breeding programs and identification of cultivars optimized for changing climate conditions.

Introduction

Plants occupy a diverse array of habitats due to key adaptations acquired in evolutionary responses to natural selection. Local adaptation is driven by divergent selection in contrasting environments, and is characterized by native populations consistently having higher fitness at their location than foreign transplants from different habitats or locations. Local adaptation can be constrained or confounded by gene flow, lack of genetic variation, genetic drift, and the genetic architecture of traits (Hereford, 2009; Kawecki & Ebert, 2004). The vast majority of plants are found to be locally adapted based on empirical studies of fitness responses in reciprocal transplant studies (Hereford, 2009; Leimu & Fischer, 2008). Yet, we still know very little about the genetic architecture underlying locally adapted traits and its role in shaping how an organism responds to the environment (Savolainen et al., 2013; Tiffin & Ross-Ibarra, 2014).

Genetic architecture is a broad concept that includes, but is not limited to, the number, magnitude and physical distribution of genetic loci that contribute to a phenotype, genetic correlations, and the direction and mode of allelic effects at those loci. Differences in these elements of a quantitative trait can shed light on both how that trait has evolved and how it may respond to evolutionary forces, particularly selection, in the future. For example, if variation in an adaptive trait is controlled by an allele at a single locus of major effect it may fix rapidly in large populations as opposed to alleles of small effect. Similarly, genetically correlated traits can result from linkage disequilibrium, physical linkage, or pleiotropic interactions (Sinervo & Svensson, 2002). Trait correlations can facilitate or constrain adaptation (Lande & Arnold, 1983). Pleiotropy

specifically may lead to trade-offs, which constrain response to selection. Context dependent effects, such as gene-by-gene interactions (epistasis) and gene by environment interactions (G x E), potentially maintain variation in a population and allow it to respond to multiple environments (Mackay et al., 2009). Understanding the genetic basis of local adaptation is not only important for evolutionary biology, but has relevance for breeding regionally adapted crop cultivars (Takeda & Matsuoka, 2008) and predicting plant responses to climate change (Reusch & Wood, 2007).

Over time, local adaptation to different habitats can contribute to the formation of distinct ecotypes within species. Ecotypes generally differ in suites of locally adapted traits from other populations (Lowry, 2012). In plants, ecotype formation is often driven by divergence in soil water availability across habitats. Ecotypes adapted to wet habitats typically are larger in size, produce more branches, and flower later than ecotypes from drier habits. Ecotype divergence has the potential to lead to speciation through ecogeographical isolation (Glennon et al., 2012; Kay, 2006; Lowry et al., 2008; Ramsey et al., 2003) but if there is no reproductive isolation, the divergent effects can be reversed when environmental selection pressures change (Gow et al., 2006; Seehausen et al., 2008; Taylor et al., 2006). Ecotypes have several features that make them ideal systems to study the genetics of local adaptation. They are not reproductively isolated so viable offspring can be generated and therefore allow a variety of genetic analyses. There are usually multiple populations within each ecotype, which can provide inherent natural replication and comparative analyses. Ecotypes are often comprised of suites of putatively adaptive traits, which facilitate the exploration of genetic correlations and pleiotropy and their impact on evolutionary responses. One interesting aspect of genetic architecture that can be addressed concerns fixed differences of allelic effects between ecotypes. An excess of large-effect fixed differences may indicate reduced gene flow or longstanding divergent selection and may give insight into how divergent traits and trait correlations have evolved (Orr, 1998; Yeaman & Whitlock, 2011). Many small effect fixed differences may be due to a large amount of standing genetic variation prior the divergent selection pressure (Barrett & Schluter, 2008; Colosimo et al., 2005). A lack of fixed differences

between ecotypes suggests adaptations may have arisen independently and in parallel across ecotype populations.

Panicum virgatum (switchgrass) is an ideal system to study ecotype divergence. Switchgrass is a long-lived outcrossing C4 perennial grass native to a large region of central and eastern North America and extending south into Central America. It is a dominant species of the tallgrass prairie, utilized as a forage crop, and a designated bioenergy feedstock (M. Casler, 2012; Michael D Casler et al., 2011; Parrish et al., 2012). Classic research in switchgrass has demonstrated that phenological traits, including date of emergence, flowering time, and date of senescence are strongly correlated with latitude of origin in common garden experiments (McMillan, 1959, 1965, 1967). Transplantation experiments and field trials have consistently demonstrated that moving genotypes north and south of their locations of origin results in a loss of fitness due to a suite of environmental factors (M. Casler & Vogel, 2004; Lowry et al., 2014; McMillan, 1959, 1965, 1967; C. L. J. Porter, 1966). Southern populations of *P. virgatum* are typically more tolerant to hot summer conditions, while northern populations are less likely to be killed by winter cold (M. Casler & Vogel, 2004). In addition to field trials, genomic tools are in development (reviewed in: Michael D Casler et al., 2011) and a reference genome has been released (*Panicum virgatum* v1.0, DOE-JGI, http://phytozome.jgi.doe.gov).

Switchgrass phenotypic diversity can be characterized into two main ecotypes, northern "Upland" and southern "Lowland", that are hypothesized to have descended from glacial refugia (McMillan, 1959; Y. Zhang et al., 2011b). Molecular studies support this by estimating deep divergence between distinct upland and lowland haplotypes to be 0.7-1 million years (Morris et al., 2011; Y. Zhang et al., 2011b). Upland varieties are adapted to colder climates, drier soils, shorter growing seasons, and range across cold hardiness zones 2-7 (USDA Zone Map; http://planthardiness.ars.usda.gov/). Lowland plants tend to be larger in growth habit, more erect in stature, and have blue-green waxy leaves. They are commonly found in riparian habitats, and occur across cold hardiness zones 6-10. Genetic variation in the cytoplasm is highly correlated with ecotype (Morris

et al., 2011; Y. Zhang et al., 2011b). Uplands can vary in ploidy from 4x-8x whereas lowlands are only 4x (though see: Y. Zhang et al., 2011b). Although the species is polyploid, recent full-sib linkage studies indicate tetraploid switchgrass maintains disomic inheritance (Lu et al., 2013; Okada et al., 2010). Tetraploids of each ecotype are largely reproductively compatible (McLaughlin & Kszos, 2005) and hybrids are found in regions of co-occurrence (Lowry et al., 2014; Y. Zhang et al., 2011a). Porter (1966) showed that upland and lowland ecotypes are locally adapted to their respective habitats through a reciprocal transplant experiment. The upland ecotype was found to be more drought tolerant and have higher nitrogen demand than the lowland ecotype, which is more tolerant to flooding. Numerous other transplant and field trials have demonstrated phenotypic and physiological differences between the two ecotypes (Barney et al., 2009; M. Casler & Vogel, 2004; Cassida et al., 2005a, 2005b; Cortese et al., 2010; Wullschleger et al., 1996; Yang et al., 2009). Another interesting divergent trait is pathogen resistance. *Puccinia emaculata* is a rust fungus that infects switchgrass. Resistance to rust is found to be heritable and southern populations are typically more resistant than northern populations (Eberhart & Newell, 1959; Uppalapati et al., 2013). Fungal pathogens are generally moisture and temperature sensitive and often require high relative humidity for infection and sporulation (Harvell et al., 2002), which may have driven greater resistance in southern populations.

We developed a new outbred genetic mapping population to understand the genetic basis of adaptation to environmental factors that are divergent between ecotypes of switchgrass. The mapping population was formed through reciprocal crosses between four grandparents derived from different locations across the Great Plains of North America. Two of the grandparents are lowland cultivars derived from the southern Great Plains, while the other two grandparents are upland cultivars from the northern Great Plains (Figure 2.1). This balanced design, including upland/lowland cytoplasm, allows us to ask whether a shared set of loci are involved in adaptive divergence between southern lowland and northern upland populations or if different alleles and loci might be involved in reaching similar phenotypes in different local ecotype populations. We assessed this

complexity through genetic mapping and characterizing allelic effects for quantitative trait loci (QTL) associated with ecotype divergence. Our results provide insight into the underlying genetic basis of how plants adapt to different environments and contribute to the switchgrass community by providing a genetic map, inclusive of ecotype variation, and relevant loci for marker assisted selection.

Methods

Outbred mapping population

P. virgatum is an obligate outcrosser (Martínez-Reyna & Vogel, 2002) and as such it is necessary to account for the fact that parental material will be genetically heterogeneous and thus generate many marker segregation types. We created a 4-way phase-known (pseudo-testcross) population to evaluate the genetic architecture of upland/lowland traits in switchgrass. In this scheme, two sets of grandparents [$lowland_1$ x $upland_2$ & $upland_3$ x $lowland_4$] were crossed to create F_1 hybrids that were then reciprocally crossed to generate two large "outbred F_2" populations ($F_{12}♀xF_{34}♂$, $F_{12}♂xF_{34}♀$) of 200 progeny each. We used four tetraploid grandparents in this design: Alamo ("AP13" genotype, the reference genome accession, a southern Texas local ecotype), West Bee Cave ("WBC3" genotype, a central Texas lowland ecotype), Summer ("VS16" genotype, a northern upland accession), and Dacotah ("DAC6" genotype, a northern upland accession). The Alamo and Summer grandparents functioned as pollen donors in crosses and therefore the two F_1 hybrids and their subsequent outbred families differ in that they contain either lowland WBC3 or upland DAC6 cytotypes (Figure 2.1). Each outbred family can segregate up to four unique alleles donated by the grandparents. Phase can be resolved from the multi-generational information. Sampling multiple grandparental alleles increases the possibility of evaluating informative QTL and inspection of the allelic effects of QTL may provide insight into genetic heterogeneity in

ecotype divergence. Finally, the cytoplasmic segregation also allows for investigation into cytoplasmic by nuclear interactions.

Cultivation and Phenotyping

Seed from the reciprocal F_1 hybrid cross was germinated and initially planted in 4-inch pots in a greenhouse at the University of Texas at Austin. At the two true leaf stage they were transferred to 1-gallon pots using a potting mix of Promix (Premier Tech Horticulture, Riviére-du-Loup, Quèbec, Canada) and Turface (Profile Products, Buffalo Grove, IL, USA) in a ratio of 4:1. Plants were grown in a greenhouse under 16h days from January to June 2012. Each plant was scored for date to first flowering, from time of transfer to the 1-gallon pot to anthesis, and height, as total length of tallest tiller on the day of first flower. Leaf tissue for genomic DNA extraction was also collected and frozen in liquid N_2. After flowering, the pots were moved from the greenhouse to an outdoor field nursery. At the end of the growing season the plants were clonally divided into two replicates and repotted in the same soil mixture using 1-gallon pots and maintained in the field nursery. One replicate of each of the mapping progeny genotypes and 20 replicates of each grandparent and F_1 hybrid were transplanted into the field at the experimental garden site at the Brackenridge Field Labs in Austin, TX in February 2014. The field planting was based on a randomized honeycomb design with 1.25 m interplant distances surrounded by a row of border plants to minimize edge effects. Weed barrier cloth (Sunbelt 3.2oz., Dewitt, Sikeston, MO, USA) was used to aid in establishment and minimize plot maintenance. The common garden field site was located in lowland switchgrass habitat associated with Colorado River floodplain (30.284138° N, -97.781632° W), where the soil type is Yazoo sandy loam.

The following traits were measured during the 2014 growing season for field grown plants: date to first flowering, chlorophyll content, specific leaf area, midday water potential, total number of tillers, single tiller mass, number of leaves on 5 total tillers, height, and pathogen susceptibility. Date to first flowering, hereafter, flowering date, is

calculated as the number of days from emergence to first anthesis. We measured midday water potential (MD WP) using a Scholander pressure chamber (PMS Instrument Company, Albany, OR, USA) on a mature leaf. Water potential indicates the leaf water potential status, and when measured during the heat of the day, can indicate tolerance to water stress. Larger (more negative) water potential values indicate more water stress (Pérez-Harguindeguy et al., 2013). Relative chlorophyll content was estimated using a chlorophyll SPAD meter (Konica-Minolta SPAD 502; Konica-Minolta, Chiyoda, Tokyo, Japan). Three readings from a single mature leaf were taken and the mean value recorded. Relative chlorophyll content, or leaf "greenness", was estimated by a corrected ratio of transmitted light with wavelengths 940 and 650 nm (Markwell et al., 1995). Specific leaf area (SLA) is the ratio of leaf area (mm^2) to dry leaf mass (g), the resulting value is a measure of leaf density (Pérez-Harguindeguy et al., 2013). We recorded the area of three mature leaves using a portable leaf area meter (LI-3000A, Li-Cor, Lincoln, NE, USA) and recorded the mass of the same leaves after full desiccation. Larger values for the area to mass ratio indicate a thinner, less dense leaf. Total plant height, single tiller mass, total number of tillers, and number of leaves on 5 tillers were taken at the end of the growing season. Height was measured using a graded measuring rod and tiller mass was calculated by averaging the dry weight of 5 mature tillers stripped of leaves and inflorescences.

Scoring for Puccinia emaculata

The fungal rust pathogen *Puccinia emaculata* was observed in the mapping population during the potted growing season in 2013. Plants were treated with Daconil (GardenTech, Palatine, IL, USA) and all but 12 inches of above ground biomass was removed before dormancy. The plants were then planted in the field for the following 2014 growing season. Physiological phenotypes (see above) were measured on green, healthy leaves. At the end of the growing season, 5 mature tillers were harvested and manually stripped of their leaves. Each leaf was then scored for presence of the fungal

pathogen using a qualitative 4 point rating, 1- completely infected and brown in color, to 4 - no evidence of pathogen and green in color. The total number of leaves on all 5 tillers was recorded as well. Each plant was then given a percentage score for each leaf category. For example, a highly susceptible plant might exhibit 78% of its leaves receiving a pathogen score of 1, 18% of its leaves a score of 2, and 5% a score of 3. We then used a principal component analysis for the four percentile scores to find the major axis of variation for the trait. The first principal component explained 86.4% of the variation and was largely associated with the percentage score of completely infected leaves. We therefore used the scores for the first principal component (Rust PC1) as a proxy for a pathogen resistance phenotype.

Genotyping

DNA extraction

Fresh leaf tissue was collected in the greenhouse, immediately frozen in liquid N_2, and stored at -80C in the spring of 2013. The equivalent of 100mg of wet tissue was then used for genomic DNA extraction with the MasterPure Plant Leaf DNA Purification Kit (Epicentre, Madison, WI, USA). The procedure was modified with an initial RNase A treatment and two subsequent EtOH washes. Final DNA was eluted in 30µl TE buffer and quantified using the Broad Range spectrum kit of the Qubit 2.0 (Life Technologies). Each extraction yielded approximately 30-200ng/µl DNA.

RAD marker library prep and Genotyping

Each plant was genotyped using a double-digest Restriction-site Associated DNA sequencing (ddRADseq) (Peterson et al., 2012) scheme. In brief, 300ng of DNA was cut with EcoRI and SphI enzymes. In-line barcodes were ligated onto the EcoRI cutsite, fragments were size selected on a Pippen Prep (Sage Science, Inc. Beverly, MA, USA) at

a 300bp +/- 30bp range, then Illumina adaptors were ligated onto the SphI cutsite. This produced 9 multiplexed libraries with 48 individuals each. Libraries were sequenced on an Illumina 2000 HiSeq (San Diego, CA, USA) at the Genome Sequencing and Analysis Facility in Austin, TX. Raw read quality was evaluated with FastQC (v. 0.10.1). Library demultiplexing was performed in Stacks (v. 1.06, Catchen et al., 2011). Sequencing resulted in ~1.6 million (mean = 1,626,329 ± 34,002.78 SE) raw reads per individual. Reads from each individual were then mapped to the *P. virgatum* v1.1 genome (DOE-JGI, http://phytozome.jgi.doe.gov) using BWA *mem* (BWA v. 0.7.9a, Li 2013; Li and Durbin 2010). Mapped reads were further processed in Stacks for genotype calling and allele designations. The two F_1 hybrids were designated as parents in the Stacks ref_map.pl pipeline and 140,561 markers were included in the catalog of markers for scoring genotypes. Each individual was then genotyped at ~ 38,000 (mean = 38,161.2 ± 670.0 SE) markers. The coverage for each marker was approximately 40x. Due to missing data the dataset was filtered for markers present in at least 50% of the individuals. We also filtered for segregation distortion by testing for significant deviation from Mendelian expectations and removing all loci where $P<0.00005$. This resulted in a high quality set of 1,348 markers available for subsequent linkage analysis.

Linkage Map assembly and QTL analysis

Outbred mapping populations are often analyzed in a pseudo-testcross approach, developing independent maternal and paternal linkage maps based on single dose markers that uniquely segregate in either parent (Haley et al., 1994). Here, we develop a joint linkage map based on the outbred full-sib family (CP) design using Joinmap (v. 4.1) and the multipoint maximum likelihood (ML) algorithm (van Ooijen, 2011). This algorithm is unique in that it simultaneously estimates phase and recombination fraction and can utilize all possible segregation types in ordering markers and estimating marker interval distances. We first grouped markers into linkage groups using a conservative logarithmic odds-ratio (LOD) of 10.0. We then used a simple regression algorithm to

order markers on each group with the following settings: pairwise recombination frequency < 0.4, LOD > 3, with a Kosambi mapping function to calculate genetic distances. Markers that affected goodness of fit (mean chi-square >3) were removed. We then used ML to order and estimate mapping distances for the edited groups. This produced a finalized map with 1,281 markers.

We mapped QTL for phenotypic traits for the upland/lowland outbred mapping population using a stepwise multiple-QTL model fitting method as implemented in R/qtl (Broman et al., 2003; Manichaikul et al., 2009). A generalized linear model was used to correct for the effects of potting cohort for traits measured in the greenhouse. Field traits did not need a cohort correction and were quantile normalized to create a uniform set of normally distributed traits. All QTL scans were performed using a normal model and Haley-Knott regression based on dense 2 cM grid of pseudomarkers generated using the calc.genprob function. We calculated LOD penalties for main effects and interactions for each trait through 1000 permutations of the scantwo function at an α of 0.05. We conducted a forward/backward stepwise search for models with a maximum of 10 QTLs that optimized the penalized LOD score criterion. We calculated the 1.5 LOD drop interval of QTLs in the best-fit models and the percent variance explained for each QTL based on the final best-fit models using the fitqtl function. We designated cytoplasm (cross direction) as an additive and interactive covariate for the scantwo penalty calculation. We also designated cytoplasm as an additive covariate for the stepwise model fitting and performed a post-hoc nuclear by cytoplasm interaction test. All computation was performed on the Lonestar cluster at the Texas Advanced Computing Center at UT Austin.

Results

Phenotypic variation, divergence and correlation

The distinct upland and lowland ecotypes of switchgrass are characterized by divergence at a variety of traits including differences in flowering time, growth architecture, physiological characteristics, and disease susceptibility (Michael D Casler et al., 2011; McMillan, 1964, 1965, 1967; Uppalapati et al., 2013). Most traits we measured showed the expected trait differentiation between the upland (U) and lowland (L) ecotypes based on the mean values of each of the grandparents including: height (L>U), tiller mass (L>U), specific leaf area (U>L), rust resistance (L>U), flowering time (L>U) and leaf number (L>U). SPAD and midday water potential did not differ between upland and lowland ecotypes, and tiller count was in the opposite direction of the pattern anticipated – lowland genotypes developed more tillers than upland genotypes during this establishment year. We found no apparent transgressive segregation of these traits in F_1 hybrids, in that all F_1 mean trait values fell within the range of grandparental values (Figure 2.2). The phenotypic distribution of the traits in the mapping progeny was generally unimodal and exhibited limited transgressive segregation (Figure 2.2). For example, flowering time in the F_2 progeny ranged from 5 to 78 days, surpassing the maximum grandparental ecotype value of 64 days from lowland AP13. We did not observe a significant effect of cytoplasm on any of the measured phenotypes (ANOVA $p > 0.05$ for all trait comparisons across cytoplasmic backgrounds, data not shown).

Ecotypes are comprised of suites of divergent phenotypes associated with adaptation to contrasting habitats. To explore patterns of trait correlation, we calculated phenotypic correlations among traits for the grandparent ecotypes and recombinant mapping progeny using the Spearman rank method and a Holm-Bonferroni multiple test correction on raw phenotype values. In the grandparents, we expected traits to correlate with each other along the expected ecotype lines and this is largely what we found, with significant correlations for 25 out of 36 pairwise trait comparisons (Table 2.1). The strongest correlations discovered were between height and tiller characteristics. All correlations were positive except with specific leaf area, where the correlations were strongly negative. This is expected as the smaller, thinner upland leaves have a larger leaf area to leaf mass ratio. The weakest correlations were with midday water potential, the

trait in which we observed the least variation among grandparental genotypes. Unlike the distinct upland and lowland grandparents, we did not observe strong trait correlations in the mapping progeny, where the number of significant correlations decreased to 14 (Table 2.1). This provided evidence that genetic recombination reduced the number of pairwise trait covariances and the traits are largely polygenic, not controlled by few pleiotropic loci of major effect.

Linkage Map

We successfully placed 1,281 ddRAD markers into 18 distinct linkage groups corresponding to the 9 chromosomes of the tetraploid switchgrass sub-genomes (Figure 2.3). The total map distance is 2,288.7 cM and the average inter-marker map distance is 1.8 cM (+/- 0.55 SE). This map distance is comparable to the "Kanlow" by "Alamo" pseudo-testcross map at 2,200.4 cM (Lowry, Taylor, et al., 2015) and larger than the "Alamo" male (1,515 cM) and "Kanlow" female (1,935 cM) (Okada et al., 2010) and the "Alamo" female (1,733 cM) and "Summer" male (1,508 cM) maps (D. Serba et al., 2013). One third of the ddRAD markers mapped to unanchored contigs in the *P. virgatum* reference genome, and we were able to place these markers into linkage groups. There are several marker segregation types that result from an outbred cross depending on the heterozygosity in each of 4 grandparents (Wu et al., 2010). For example, 4 unique grandparental alleles <abxcd> result in a fully informative marker for both linkage and QTL mapping. However loci with fewer than 4 unique alleles result in a partially informative marker (e.g., <efxeg>). We found 798 partially informative bi-allelic markers, 450 partially informative tri-allelic markers and 33 fully informative markers. The fully informative markers uniquely identify the contributed grandparental chromosomes and were used to phase each linkage group for analysis of allelic effects.

QTL

Overall, we identified 33 QTL and 3 epistatic interactions across 11 traits using stepwise model selection (Figure 2.3). The largest additive effect QTL was located on linkage group (LG) 5a with 19.3 percent variance explained (PVE) for flowering time in the greenhouse. The largest additive effect QTL in the field were for specific leaf area on LG 9a (13.46 PVE) and plant height on LG 2b (13.34 PVE) (Table 2.2). The only trait lacking significant QTL was midday water potential. Notably, we found a total of 7 QTL on LG 9b. Over two thirds of the QTL peaks were found on one third of the linkage groups, with 72.7% of the QTL localized to the tetraploid homeologue pairs 2, 5 and 9. We did not detect any significant nuclear by cytoplasmic interactions.

We found 6 QTL and 1 epistatic interaction for specific leaf area, the largest number of QTL per trait in our study. Together, these QTL totaled 57.23 PVE for that trait. We also found 5 QTL for both measurements of height, in the greenhouse and in the field. Interestingly, there was only one location, on LG 9b, where QTL for these traits colocalized. We detected 3 and 4 QTL for flowering time in the greenhouse and in the field respectively. However, while they each had a QTL on LG 2b, the confidence intervals did not overlap. We discovered 3 pairwise epistatic interactions that explained a moderate amount of variance relative to the additive effects. These interactions were found for height (8.12 PVE), specific leaf area (9.43 PVE) and greenhouse flowering time (11.22 PVE). We found significant rank changes in the allelic effects depending on genetic background within each case but there was no consistent pattern across the 3 separate interactions. Three traits, leaf number, SPAD, and rust resistance, yielded a single QTL but each QTL was less than 10 PVE, suggesting there may be many undetected loci controlling these traits.

Allelic Effects

The use of an outbred mapping population affords a more detailed evaluation of QTL allelic effects than traditional inbred line crosses given the contribution of 4 grandparental alleles in the F_2 progeny. After phasing, the allelic affects at QTL can be

evaluated based on the 4 possible recombinant progeny including upland homozygotes (VS16/DAC6), lowland homozygotes (AP13/WBC3), and the two upland/lowland hybrid genotypes (VS16/AP13; DAC6/WBC3). Inspection of the additive effects can therefore allow some interpretation of the distribution of functional alleles within and between ecotypes and the genetic mechanism associated with each trait. For example, the functional alleles can be consistent between ecotypes where unique upland and lowland genotypes contribute the same functional alleles. Or functional alleles can be polymorphic within ecotypes where individual genotypes contribute unique functional alleles to progeny irrespective of ecotype, or in our case, only observed in one side of the cross design. The direction of allelic effects and patterns of dominance can also shed light on genetic mechanisms.

We were able to categorize 26 of the QTL allelic effects as either replicated or non-replicated within each ecotype; where replicated effects suggest fixed differences between ecotypes. We found 15 QTL with replicated ecotypic divergence suggestive of an intercross configuration (U/L x U/L) generating upland/upland homozygotes, lowland/lowland homozygotes, and upland/lowland heterozygote genotypes (Table 2.2). Replicated divergent effects include several allelic effect patterns: for example, both upland/lowland heterozygous genotypes were an intermediate value between the homozygous upland/upland and lowland/lowland genotypes (Figure 2.4a-b). We found 9 of the replicated divergent effects indicated dominance, where a single homozygous lowland/lowland (or upland/upland) genotype differed from the other three genotypes with at least one upland (or lowland) allele (Figure 2.4c). And lastly we identified cases where all alleles were additive in the same direction, as replicated divergence because the effects are consistent across all 4 grandparental genotypes (Figure 2.4d). Conversely, the non-replicated cases suggest functional variation between populations rather than between ecotypes. We found 11 polymorphic effects, with 6 in the AP13 x DAC6 cross and 4 in the WBC3 x VS16 cross and one that was indistinguishable between the two (Figure 2.4e-f, Table 2.2).

In addition to the patterns of allelic effect, we were interested in the direction of QTL effects in relation to observed ecotype divergence. We found 10 of the 33 QTL had additive effects in the opposite direction than expected for divergence between upland and lowland ecotypes (Table 2.2). For example, lowland ecotypes consistently flower later than uplands, but the lowland alleles for flowering date on LG 2b resulted in an earlier flowering time than upland alleles (Figure 2.4b). All 5 QTL on LG 2b, 3 out of 4 QTL on LG 5a, and one each on LGs 9a and 6b exhibited allelic effects in the opposite of expected upland/lowland ecotypic divergence. We found 4 of the inconsistent QTL effects were split between replicated and non-replicated allelic effects types and 6 inconsistent QTL did not display a clear allelic effect signal.

Discussion

We assembled a genetic linkage map from a reciprocal 4-way cross for switchgrass, mapped relevant QTL, and characterized functional allelic effects for traits associated with ecotype divergence. We were able to successfully recombine upland and lowland alleles as evidenced by the reduction in significant trait correlations from grandparents to mapping progeny. Our linkage map is novel in that it utilizes the latest genotyping technology and integrates a substantial spectrum of variation within and between the upland and lowland ecotypes into a single consensus map. This allowed us to characterize the genetic architecture of ecologically divergent traits that may shed light on potential mechanisms of ecological divergence between upland and lowland switchgrass.

Genetic architecture

The genetic architecture of adaptive divergence has not been well characterized, largely due to the complex polygenic nature of most ecologically important traits and a

poor understanding of the link between genotype and phenotype (Bergelson & Roux, 2010). We found that all traits were polygenic and we did not find evidence of loci of major effect. The greatest number of QTL detected for a single trait was 6 (SLA) but we can assume that the majority of small effect loci are below the level of detection. We found that ecotypic divergent traits are composed of mixed allelic effect types and directions. We did not find any major patterns underlying the genetic architecture of these traits but instead found evidence of several factors that can lead to limited or enhanced response to selection such as linkage disequilibrium (LD), pleiotropy, or physical linkage, epistasis, and dominance. Thus the genetic patterns of divergence are unique for each trait and should be considered carefully for breeding and conservation efforts.

An open question is the degree to which adaptive suites of traits are the result of pleiotropic gene action, physical linkage, or LD generated by strong divergent selection on ecotypes. In a scenario where there is little to no gene flow between ecotypes, and selection is acting on suites of beneficial traits, we expect LD to build up between loci associated with all traits that contribute to the locally adapted phenotype. We also expect to see a significant reduction in LD and subsequent correlational structure following random mating in our mapping study. In our mapping progeny we observed that 11 of the trait correlations that were significant in the grandparents were no longer significant in the mapping progeny, suggesting that they were associated through LD. Traits can also be correlated through physical linkage or pleiotropy. For both of these possibilities we expect traits to remain significantly correlated in the mapping progeny as well as share colocalized QTL. If traits are strongly correlated after a few rounds of random mating, have colocalized loci, and the allelic effects are in line with ecotype divergence, this will not constrain local adaptation. However, if allelic effects are in the opposite direction or if the trait correlation is negative with respect to the adaptive suite of traits associated with each ecotype, then response to selection will be constrained due to genetic trade-offs. For example, consider two trait pairs: tiller mass and SLA, and height and SLA, that were significantly correlated in both the grandparents and the mapping progeny. The raw correlation value was negative but this is the expected direction for ecotype divergence.

Tiller mass and SLA shared overlapping QTL confidence intervals on LG 9b and both had effects in line with ecotypic divergence. In contrast, height and SLA on LG 5a differed in direction of allelic effects. The SLA allele confounded ecotypic divergence and height contributed to it. This will constrain response to selection. However, the size of the effects dictates the magnitude to which the linked traits will constrain response to selection.

Our results are consistent with other *P. virgatum* studies and we provide possible evidence for genomic regions across the *Panicum* complex with consistent and conserved effects on traits. We found several regions in the genome where QTL confidence intervals overlapped across traits, mainly on tetraploid homeologous pairs 2, 5, and 9. LG 2b is particularly interesting because all of the QTL effects were in the opposite direction than expected based on patterns of ecotypic divergence. This may be due to several factors that we address below. We found 7 QTL that localized to LG 9b. This is consistent with previous mapping efforts of similar traits by Serba et al. (2014) and Lowry et al. (2015). Both studies found major biomass QTL on LG 9b. Another interesting comparison is to QTL found in diploid *Panicum hallii*. The reference genome for switchgrass is anchored on the *P. hallii* genome and switchgrass and *P. hallii* are estimated to be 5.3 million years diverged (Y. Zhang et al., 2011b). *P. hallii* has two genomic hotspots of ecotype divergence, where QTL colocalized on LG 3 and 5 (Lowry, Hernandez, et al., 2015). LG 5 is particularly interesting because QTL for tiller number, flowering time and height colocalized to the lower arm of LG 5 in *P. hallii* and LG 5a in switchgrass. Additionally, the lower arm for switchgrass LG 5b contained greenhouse-specific QTL for flowering time and height. Lastly, the earliest polyploidization in switchgrass was estimated to be around the same time as ecotype divergence ~ 1.2 mya (Y. Zhang et al., 2011a). We found that specific leaf area had two QTL on the lower arms of LG 5a and 5b and the upper arms of LG 9a and 9b suggesting possible sub-functionalization of loci rather than pseudogenization as a result of genome duplication.

The last component of genetic architecture that we address is context dependent effects. Gene by gene or epistatic interactions can be difficult to detect because the

sample size of two-locus individual comparisons is reduced compared to the single-locus individual comparisons that are tested for additive effects (Mackay et al., 2009). In general, epistatic effects can select against distant crosses that may break apart beneficial interactions and introduce deleterious effects. We found three epistatic effects that account for a significant amount of variation in each trait. This may be a mechanism for ecotype divergence. Interestingly, cytotype is strongly divergent between upland and lowland (Morris et al., 2011; Y. Zhang et al., 2011b) but we did not find significant cytotype by QTL effects for the traits we measured. Another major type of context dependent effect is gene by environment (G x E) interactions. A common observation is the occurrence of "variety by environment" interactions for biomass productivity in switchgrass (M. D. Casler & Boe, 2003; M. D. Casler et al., 2007; M. Casler & Vogel, 2004; e.g., A. Hopkins et al., 1995a, 1995b). For example, Casler & Boe (2003) found extensive rank changing in biomass yield and tissue characteristics among six varieties grown in two locations. We did not explicitly test for G x E in this experiment but there is precedent for environment playing a major role in switchgrass growth (M. Casler & Vogel, 2004); and putatively adaptive traits, such as those differentiating upland and lowland ecotypes, are be expected to exhibit G x E (Des Marais et al., 2013). We did however measure two traits, flowering date and height, in both a greenhouse environment during the initial growth phase, and in the field. We did not find any significant overlap in shared QTL for each trait but did find that the traits measured in the field had more allelic effects in the opposite direction of ecotypic divergence. This could simply be due to the greater environmental control in the greenhouse or the confounding factor of growth stage. But it may also partly be due to the lowland habitat of the field site. We conducted this experiment in a lowland habitat and as such the upland allelic effects may differ from those in their home environment. Or, a more complex scenario may be that both types of alleles may have epistatic interactions that we did not detect.

Allelic effects

Since upland and lowland ecotypes are estimated to have diverged ~1.3 mya, during the Pleistocene, we might expect the majority of functional alleles to be fixed between ecotypes because of putatively strong and consistent divergent selection pressures. However, if there are very low levels of gene flow across the geographic range of switchgrass populations we might expect functional alleles to be unique within each population due to convergent evolution. Not surprisingly, we found a combination of replicated and non-replicated allelic effects within the ecotypes. This is consistent with a model of ecotype divergence stemming from ancient glacial refugia followed by adaptive radiations and genetic bottlenecks with each glacial cycle (Y. Zhang et al., 2011b). Ecotypes were originally formed from strongly selected and possibly small populations. However, there have been several glacial episodes since the estimated time of ecotype divergence, and with that, additional population bottlenecks and subsequent bouts of range shifts and changes in environmental selection pressures. We did not have the statistical power to determine if specific traits are comprised only of replicated or non-replicated allelic effects but this analysis lays the groundwork for future studies in this area.

Another important aspect of characterizing QTL is the direction of allelic effects. Even though QTL studies are known to both over and under estimate effect size (Beavis, 1998) there is very little bias in detecting direction of effect. We found that 30% of our QTL had allelic effects in the opposite direction than expected between upland and lowland ecotype divergence. Here we offer some possible explanations. In addition to the directional selection imposed by heterogeneous habitats, populations may experience stabilizing selection, or maintain polymorphisms in populations in the overlapping ranges that experience higher levels of gene flow. In the case of flowering time and height, two traits that are strongly divergent between the ecotypes, we found half of the allelic effects were in the opposite direction of expected ecotype divergence. One possibility in this case is that the trait is composed primarily of many very small effect loci with allelic effects in the expected direction that are beyond the level of detection of this study. We found one particular region on the lower arm of LG 2b where all effects were in the

opposite direction of ecotypic expectation. This may be the result of a chance fixation of a maladaptive chromosome block due to a population bottleneck (Orr, 1998) or from a recent selective sweep for a trait we did not measure such as leaf color or rhizome characteristics.

One exciting applied result is the location of a rust resistance locus. It is known that lowland switchgrass genotypes are more resistant to rust than uplands (A. Hopkins et al., 1995a; Uppalapati et al., 2013). This may have to do with moisture sensitivity of fungal pathogens and the mesic lowland habitat. Whatever the cause, little is known about the genetic architecture of this resistance. The natural infection that occurred during our experiment allowed us to screen for resistance QTL. We detected one resistance allele donated from the lowland WBC3 genotype using our coarse scale phenotype method. This lays the groundwork for future pathogen mapping efforts in switchgrass that can be used to develop a panel of resistance alleles. Marker assisted breeding programs can quickly and efficiently develop cultivars for different climactic regions for traits not amenable to transgenics. Locating causal genes within QTL is costly, time intensive and not useful if the phenotypic effect of each QTL is small. Yet we can introduce numerous small effect resistance alleles into the upland genetic background through several rounds of targeted marker assisted breeding to produce large phenotypic effects (Vogel & Jung, 2001).

We have designed this study to extend into future work beyond that of ecotype divergence. Clonal replicates of this mapping population will be planted across a latitudinal gradient to address clinal variation and G x E in *P. virgatum*. These complex interactions, in addition to ecotype divergence, are important considerations in habitat restoration, plant breeding, and accuracy of agronomic modeling. In addition to a conceptual framework for genetics of locally adapted ecotypes, we provide a starting point for marker-assisted selection of desired traits in a lignocellulosic bioenergy feedstock. Understanding and exploiting locally adapted traits in different genotypes will allow us to efficiently grow switchgrass in many different geographical regions economically and with minimal input and ecological impact.

Tables

	Flower Date	Height	Tiller Mass	Tiller Number	Leaf Number	MD WP	SPAD	SLA	Rust PC1
Flower Date		0.69 ***	0.571 ***	0.57 ***	0.423 *	0.27 ns	0.551 **	-0.641 ***	0.611 ***
Height	0.265 ***		0.869 ***	0.84 ***	0.577 ***	0.293 ns	0.629 ***	-0.784 ***	0.7 ***
Tiller Mass	0.232 ***	0.639 ***		0.815 ***	0.536 **	0.179 ns	0.482 *	-0.732 ***	0.725 ***
Tiller Number	0.157 ns	0.341 ***	0.165 *		0.538 **	0.25 ns	0.602 ***	-0.635 ***	0.582 ***
Leaf Number	0.066 ns	0.314 ***	0.408 ***	0.066 ns		0.233 ns	0.379 ns	-0.544 **	0.317 ns
MD WP	0.216 **	0.09 ns	0.09 ns	0.064 ns	0.003 ns		0.368 ns	-0.289 ns	0.233 ns
SPAD	0.101 ns	0.14 ns	0.151 ns	0.1 ns	-0.041 ns	0.084 ns		-0.463 *	0.413 ns
SLA	-0.124 ns	-0.22 ***	-0.271 ***	0.065 ns	-0.199 **	-0.262 ***	-0.04 ns		-0.609 ***
Rust PC1	0.149 ns	0.071 ns	-0.009 ns	-0.232 ***	-0.087 ns	0.221 ***	0.053 ns	-0.147 ns	

Table 2.1 Phenotypic trait correlations

Spearman's rank correlation (rho) for phenotypic traits within the four grandparent lines (upper-triangle) and mapping progeny individuals (lower-triangle) measured in the field, *p<0.05, ** p<0.01, *** p<0.001. P-values corrected for multiple tests using the Holm-Bonferroni method. Abbreviations: midday water potential (MD WP), chlorophyll content (SPAD), specific leaf area (SLA), pathogen resistance (Rust PC1), not significant (ns).

Phenotype	Linkage Group	Position	1.5 LOD CI	LOD	PVE	Effect Direction	Effect Type
Flowering Date	2b	113.28	106-122	9.13	9.1	+	P1
Flowering Date	4a	83.41	54-86	8.77	8.72	-	R
Flowering Date	5a	98.4	86-172	5.32	5.17	+	P2
Flowering Date	9a	132.87	16-158	5.09	4.94	-	RD
Height*	2b	132.27	126.14-132.27	14.01	13.34	+	RD
Height	3b	106.5	95.3-122	5.68	5.13	-	NA
Height	5a	150	130-181.56	4.39	3.93	+	RD
Height*	6b	61.39	56-65.35	13.32	12.63	-	P
Height	9a	130	120.45-142.6	5.81	5.25	-	P2
Tiller Mass	2b	132	120-132.27	6.35	6.48	-	NA
Tiller Mass	9b	24	10-49	6.55	6.69	+	R
Tiller Mass	2a	126	118-134	6.85	7.01	+	R
Tiller Mass	3b	114	96-124	7.45	7.65	+	R
Tiller Number	1a	70.15	59.25-79.25	7.1	8.19	+	RD
Tiller Number	5a	98.4	86-136	4.46	5.06	+	RD
Tiller Number	9b	83.86	76-94	4.43	5.02	-	NA
Leaf Number	9b	42.37	30-48	6.36	8.25	+	RD
SPAD	2b	113.28	96-123.5	5.67	7.43	-	NA
Specific Leaf Area	5a	142.18	133.1-181.56	4.57	3.68	-	NA
Specific Leaf Area	5b	104.11	100-110	7.4	6.06	+	P1
Specific Leaf Area*	8a	24.38	22-30	13.87	11.85	+	P2
Specific Leaf Area*	9a	36.63	34.39-40	15.59	13.46	+	RD
Specific Leaf Area	9b	28	22-51.984	8.75	7.22	+	P1
Specific Leaf Area	9b	108	48.21-138	4.64	3.73	+	RD
Rust PC1	8b	13.21	6-52	4.86	6.22	+	P2
GH Flowering Date*	2b	38.23	34-41.32	12	11.9	-	NA
GH Flowering Date*	5b	116.12	112-119.8	18.67	19.31	+	NA
GH Flowering Date	9b	98	52-114.79	4.31	4.08	+	P1
GH Height	3a	44	32.24-102.27	6.04	5.47	+	P1
GH Height	5b	119.8	110.13-122	6.08	5.5	+	R
GH Height	7a	40.69	20-54	5.06	4.54	+	RD
GH Height	9a	102	70-130	4.79	4.3	+	P1
GH Height	9b	69.18	56-86	5.41	4.88	+	R

Table 2.2 QTL for ecological traits measured in *P. virgatum*

*Trait** indicates an epistatic interaction. Genomic position and 1.5 Logarithm of Odds (LOD) confidence interval (CI) in centimorgans (cM). PVE is the percent of phenotypic variance explained by the QTL. Effect direction (+) indicates allelic effect is consistent with ecotype divergence, (-) is opposite of ecotype expectation. Effect type includes (R) replicated difference within ecotype, (RD) replicated difference with dominance, (P1) polymorphic allele in AP13 x DAC6 cross, (P2) polymorphic allele in VS16 x WBC3 cross, (P) polymorphic effect that cannot be resolved between cross, (NA) effect cannot be resolved. GH- greenhouse traits.

Figures

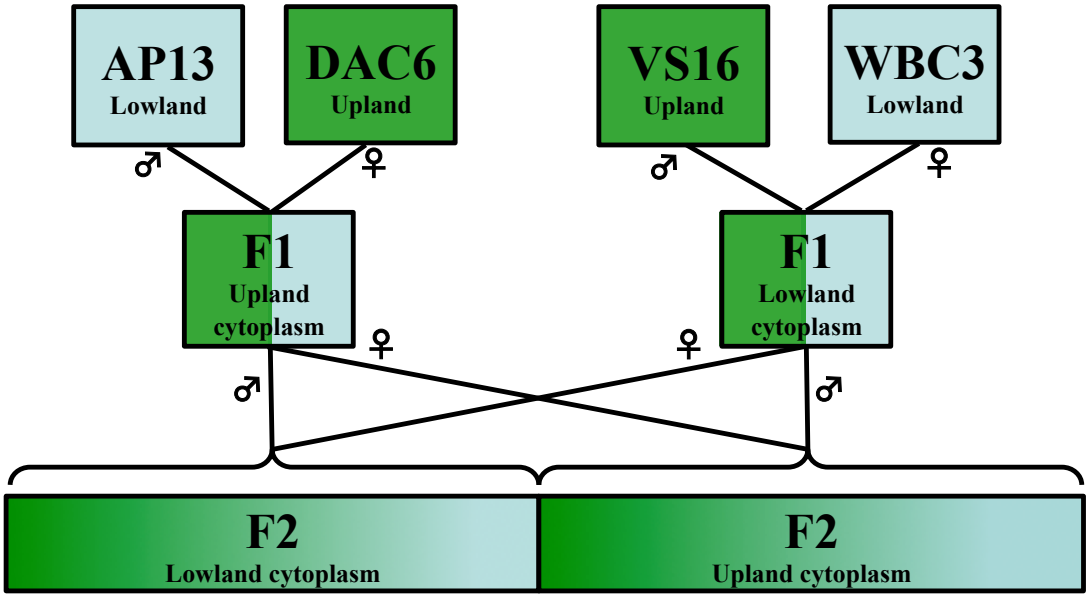

Figure 2.1 Diagram of cross

Diagram of four-way outbred reciprocal cross between two upland and two lowland ecotypes of *P. virgatum*.

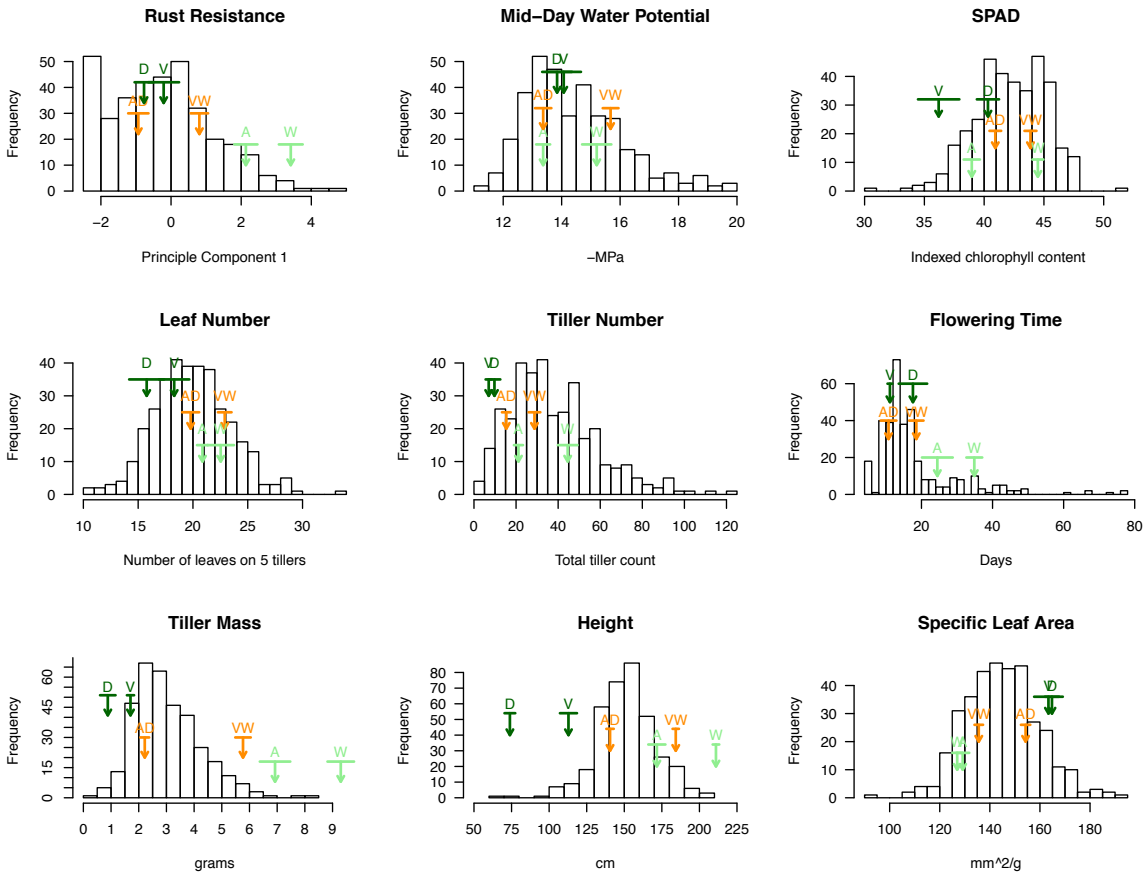

Figure 2.2 Phenotypic trait distributions for *P. virgatum* mapping population

Histograms of raw phenotypic values in the mapping population for each trait measured in the field, with mean values for each of the 2 original upland (dark green), lowland (light green) and F_1 hybrid (orange) individuals indicated by arrows, and standard error indicated by horizontal bar.

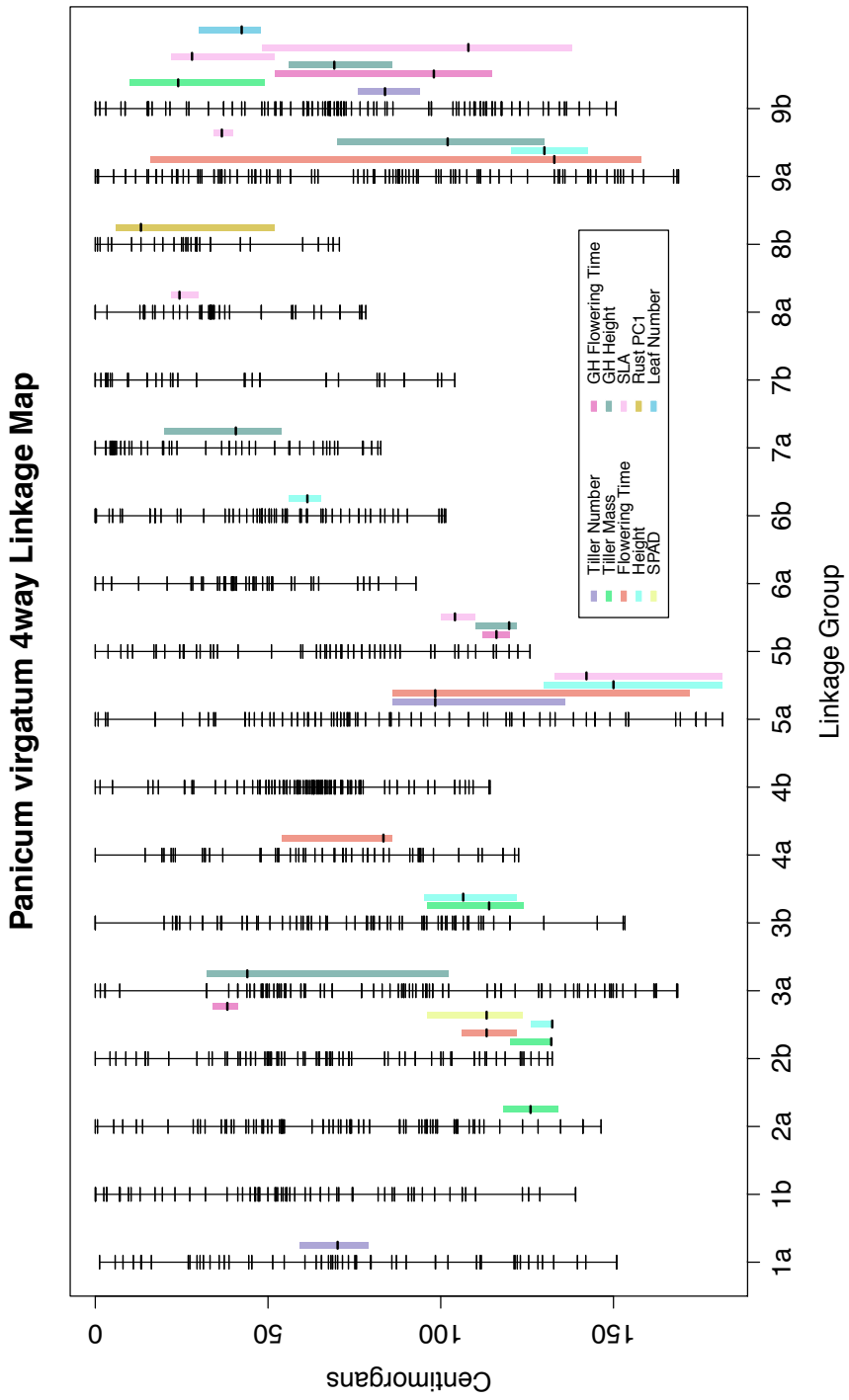

Figure 2.3 Genetic linkage map for P. virgatum with ecological QTL

1.5 LOD drop confidence intervals mapped to the right of respective linkage groups.

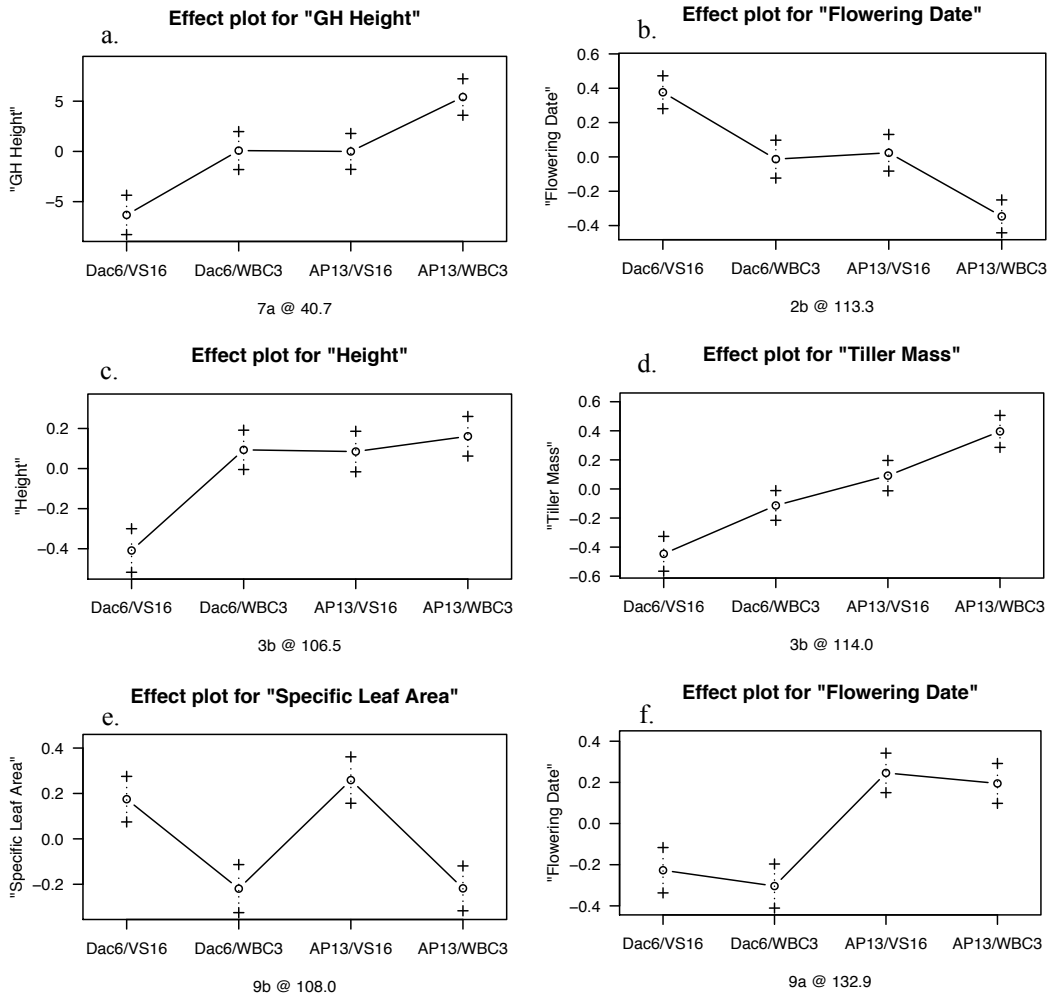

Figure 2.4 Allelic effects

Selected allelic effects plots to illustrate a) replicated ecotype differences, b) replicated ecotype differences in the opposite direction of ecotype divergence, c) polymorphic alleles in the VS16 x WBC3 cross, d) polymorphic alleles in the AP13 x DAC6 cross, e) dominance and f) pure additive allelic effects at specified loci. The x-axis indicates genotype; Dac6/VS16 = Upland/Upland , Dac6/WBC3 and AP13/VS16 = Upland/Lowland and AP13/WBC3 = Lowland/Lowland. Subtitle indicates LG and marker position of the specific locus. GH- greenhouse traits.

Chapter 3: Quantitative Trait Loci For Cell Wall Composition Traits Using Near-Infrared Spectroscopy in The Model C4 Perennial Grass *Panicum hallii*

Abstract

Second-generation biofuels derived from vegetative lignocellulosic plant material are an important component for current renewable energy strategies. Improvement efforts in lignocellulosic biofuel feedstock crops have been primarily focused on increasing biomass yield without consideration for tissue quality. Four primary components found in the plant cell wall compose the bulk of lignocellulosic plant material and contribute to the overall quality of plant tissue and biofuel convertibility characteristics. Cellulose and hemicellulose polysaccharides are the targets for fuel conversion, while lignin and ash provide structure and defense. Cell wall recalcitrance is the trade-off of cell wall composition between energy storage, structure, and defense. We explore the genetic architecture of tissue characteristics using a quantitative trait loci (QTL) mapping approach in *Panicum hallii*, a model lignocellulosic grass system. We use Near-Infrared Spectroscopy (NIRS), a fast, accurate, and inexpensive method to quantify cell wall components, and create a *P. hallii* specific calibration model. We find that ash, lignin, glucan, and xylan comprise 68% of total dry biomass in *P. hallii* and the cell wall composition is comparable to other C4 perennial lignocellulosic feedstocks. We identified 14 QTL and one epistatic interaction across these four cell wall traits and found almost half the QTL localized to a single linkage group. This is the second study to use NIRS for QTL mapping of cell wall traits. Leading lignocellulosic grasses have unwieldy genomes with varying amounts of genomic resources. As such, understanding the genetic architecture of tissue traits in a tractable model grass system will lead to a better

understanding of cell wall structure and function as well as provide genomic resources for bioenergy crop breeding programs.

Introduction

Second-generation biofuels such as ethanol, butanol, and hydrocarbons are derived from vegetative lignocellulosic plant material (Atsumi et al., 2008; Carriquiry et al., 2011; Tilman et al., 2006) and are an important component for current renewable energy strategies. These biofuels are advantageous over current first-generation grain based biofuels because they use whole plant biomass and can have reduced ecological impact on land and water resources (Sluiter et al., 2010; Somerville et al., 2010; Tilman et al., 2006). Lignocellulosic feedstocks include perennial prairie grasses such as switchgrass and big bluestem, tropical grasses such as *Miscanthus* and *Sorghum*, hardwoods such as poplar, and agricultural residues such as corn stover and sugarcane bagasse. These feedstocks have the potential to generate two to three times more biomass than first-generation grain based feedstocks (McLaughlin & Kszos, 2005; van Zyl & den Haan, 2013) annually on marginal or non-agriculture land, or as secondary agricultural products. Current second-generation fuel conversion methods estimate 70-90% recovery of glucose and other soluble carbohydrates necessary for bioethanol and other types of biofuel conversions from these feedstocks (Dien et al., 2006; Luterbacher et al., 2014).

Historic improvement efforts in lignocellulosic biofuels have primarily focused on increasing biomass feedstock yield (Bouton, 2007; McLaughlin & Kszos, 2005; Somerville et al., 2010). However, the production of quality feedstock is just as important as quantity. Decades of forage research has found that the quality of feedstock can affect the digestibility of forage in rumen guts (Jung & Allen, 1995) and can lead to increases in milk, fiber and biofuel conversion yields (Amaducci et al., 2000; Lorenz, Anex, et al., 2009; Nelson & Moser, 1994; Oba & Allen, 1999). Feedstock quality for lignocellulosic plants is dependent on the composition of the cell wall. The fuel precursor carbohydrates

in lignocellulosic feedstocks are bound in crystalized polysaccharide polymers and interwoven with a lignin matrix that both provides structure to the plant and protection from herbivores and pathogens (F. Chen & Dixon, 2007; Himmel et al., 2007). High quality biofuel feedstocks have large quantities of accessible carbohydrates while maintaining structural integrity and defense mechanisms in the field.

Cellulose, hemicellulose, and lignin are the three main components of the cell wall in lignocellulosic plants (Åman, 1993). Cellulose is a polysaccharide composed entirely of the 6-carbon monosaccharide glucose. Hemicellulose is a polysaccharide composed of a mix of 5- and 6-carbon monosaccharides with the primary component in monocotyledons being 5-carbon xylan (DeMartini et al., 2013; Dien et al., 2006). Crystalline cellulose and hemicellulose molecules are intertwined with a phenylpropanoid polymer lignin matrix and provide both structural support and defense against natural enemies (F. Chen & Dixon, 2007). Recalcitrance is the ability of the plant cell wall to resist digestion and degradation by enzymes and pathogens, though it is sometimes narrowly defined as the amount of lignin and other cell wall components that restrict access to carbohydrates in cell wall tissue. However, recalcitrance can broadly be thought of as resource trade-off between production of accessible carbohydrates for plant growth and production of structural components such as lignin. Recalcitrance is an evolved defense against herbivores and pathogens feeding on plant sugars (Buxton & Casler, 1993; Himmel et al., 2007) but it is this very phenomenon that current fuel conversion processes need to overcome to harness the energy stored in plant tissue. Therefore, it is important to understand the structure and function of the cell wall for general biological processes as well as for lignocellulosic biofuel production (Vermerris et al., 2007).

Biofuel conversion technologies are in a state of continuous development and improvement. The current conversion process of bioethanol, the most widely produced liquid biofuel from lignocellulosic biomass, starts with a mechanical or thermochemical pretreatment. This breaks down the lignin matrix and exposes the crystalline polysaccharides. Pretreatment is followed by simultaneous saccharification and

fermentation (SSF) which is a combined process of enzyme hydrolysis of structural sugars and yeast fermentation to produce alcohol (Dien et al., 2003; Sun et al., 2000). Ongoing improvements in biofuel conversion processes include a new approach that converts a small amount of the available biomass sugar into a platform molecule, gamma-valerolactone (GVL). This platform molecule acts as a solvent for the bulk of the biomass, thus avoiding the use of costly enzymes and potential tissue damaging pretreatment steps. It uses a biomass-derived product and is able to recover 70-90% solubilized sugar for downstream fermentation (Luterbacher et al., 2014). Another area of development concerns the conversion efficiency of specific yeast strains on 5- and 6-carbon sugars. The complex polysaccharides in hemicellulose can add to biofuel conversion efficiency or can detract depending on the ability of the yeast strain to ferment different types of sugar (DeMartini et al., 2013). Thus far, total conversion efficiency is a function of carbohydrate content, lignin content, inhibitory proteins, temperature, and yeast strains. Independent of the method, biofuel conversion processes will require uniform and well-defined plant tissue. Phenotypic and genotypic characterization of cell wall components and their interaction with agronomic growing conditions in the field will contribute to quality biomass production.

Plant tissue characterization in forage crops has been historically well studied in the field of agronomy and is based on a number of longstanding methods. However, some popular methods can be inaccurate or costly for large scale studies. The long used detergent analysis method provides only a coarse quantification of cell wall components and has many known biases for lignocellulosic tissue (Udén et al., 2005; Vogel & Jung, 2001; Wolfrum et al., 2009). Current procedures based on the Uppsala method (Theander et al., 1995) are more accurate but can be time and cost prohibitive for large sample studies. Near-infrared spectroscopy (NIRS) is a cheap, quick, and non-invasive method for studying cell wall components. NIRS is a two-step process that uses laboratory compositional analysis and near-infrared spectral data to build a compositional model that is used to predict sample composition based on spectral data. The laboratory component involves performing small-scale bench-top laboratory analytical procedures

(LAPs) using an updated Uppsala method (Sluiter et al., 2010) on a subset of 'calibration samples'. The spectral component involves a multivariate analysis of spectral data from the calibration dataset, which is then used to build a predictive model that can be applied to a larger spectral dataset of samples. NIRS has been used for a variety of agricultural applications from estimating seed fat content to green tea leaf alkaloids (Roberts et al., 2004). In biofuels, NIRS has been used to characterize cell wall components of switchgrass (Vogel et al., 2011), corn stover (Lorenzana et al., 2010), *Miscanthus* (Haffner et al., 2013), *Sorghum* (Guimarães et al., 2014; Wolfrum et al., 2013), mixed grasses (Payne & Wolfrum, 2015), and mixed wood (Hou & Li, 2011) among others. Calibration models are most accurate when used to predict strict tissue composition (Payne & Wolfrum, 2015) but can also include derived components such as total carbohydrate release (Lorenzana et al., 2010; Wolfrum et al., 2013) and theoretical ethanol yields (Vogel et al., 2011). Several important applications have resulted from using NIRS for rapid analysis of cell wall traits. It is easy to assess biomass quality upon arrival at a biorefinery and assess quality differences across environments, as water and other abiotic factors are known to have a large impact on yield and other biomass traits (Patakas, 2012). NIRS can aid in the refinement of conversion methods by allowing rapid analysis of different pretreatment and saccharification steps. However, one underutilized function of this high-throughput method is to understand the genetic components of biofuel cell wall traits.

Understanding the genetics of cell wall components will lead to a better understanding of cell wall recalcitrance (Bouton, 2007) as well as aid the generation of high quality feedstock. The genetic architecture of economically relevant traits is important for locating large effect functional variants in the genome and for understanding how a quantitative trait, like tissue composition, will respond to selection in breeding programs. Genetic mapping of quantitative trait loci (QTL) is the first step in locating large effect variants and determining the genetic architecture of a trait and in implementing marker assisted selection in breeding programs. Thus far, genetic analysis of cell wall traits using NIRS is limited to two studies in corn stover (Lorenz, Coors, et

al., 2009; Lorenzana et al., 2010). We are aware of only one study that maps QTL for tissue characteristics predicted by NIRS. In that study, Lorenzana et al. (2010) found significant genetic variation, moderate heritability, and many QTL with small effects. There have been a number of QTL studies for tissue characterization in forage crops based on the detergent system of analysis (Van Soest & Wine, 1967). However, results from these analyses are known to underestimate lignin two to four fold and overestimate cellulose and hemicellulose due to incomplete solubilization of protein and other inhibitory elements in lignocellulosic tissue (Vogel & Jung, 2001). Genome wide studies like QTL mapping and genome wide association mapping (GWAS) give rapid benefits for marker-assisted breeding programs and long term benefits for fine mapping, characterization of genetic architecture, and the ultimate discovery of genes involved in important biological processes.

Hall's panic grass (*Panicum hallii* Vasey) is an important genetic model system for C4 perennial grasses and for lignocellulosic biofuel crops (Lowry et al., 2013; Lowry, Hernandez, et al., 2015; Meyer et al., 2012). *P. hallii* has a small (~550Mbp) diploid genome, small stature, short generation time, and is self compatible (Lowry et al., 2013). *P. hallii* has two distinct varieties, var. *hallii* found in xeric habitats and var. *filipes* found in mesic habitats (Figure 3.1). The varietal distinction is similar to northern upland and southern lowland ecotypes in *Panicum virgatum* (switchgrass) (Lowry, Hernandez, et al., 2015), making *P. hallii* a good ecological model as well as genetic model system. Current genomic tools include an annotated reference genome (phytozome.net), transcriptome datasets, and a number of genetic mapping resources (Lowry et al., 2013; Lowry, Hernandez, et al., 2015; Meyer et al., 2012). *P. hallii* shared a common ancestor with *P. virgatum* approximately 5 million years ago (Y. Zhang et al., 2011b). This phylogenetic proximity allows for close synteny between *P. hallii* and *P. virgatum*, while the small tractable genome and self-compatibility are useful for functional assays and laboratory functional genomic experiments.

The purpose of this study is to explore the genetic architecture of tissue characteristics in a biofuel feedstock model grass. We do so using a QTL mapping

approach in *P. hallii*. We use wet lab chemical analyses of reference samples and NIRS to build a *P. hallii* specific calibration model and quantify the cell wall composition and phenotypic correlations of tissue characteristics in an F_2 mapping population between mesic and xeric varieties. We compare the cell wall composition of calibration samples to other lignocellulosic feedstocks to evaluate *P. hallii* as a model lignocellulosic grass. Both the *P. hallii* specific NIRS calibration model and exploration into the genetics of lignocellulosic cell wall traits provide valuable resources for crop improvement in bioenergy grasses and further investigation into cell wall recalcitrance.

Methods

Plant Material and Genetic Map

The mapping population was generated by crossing single inbred accessions of two morphologically distinct varieties of *P. hallii*, where var. *hallii* (HAL2 genotype) was the dam and var. *filipes* (FIL2 genotype) was the sire. Both accessions were collected from wild populations in 2010 and the cross was made in 2011. A single self-pollinated F_1 hybrid generated all the seeds for the F_2 mapping population. The F_2 progeny were grown under 16h days in a glasshouse at the University of Texas at Austin in the fall of 2011. Details of greenhouse propagation as well as QTL for morphological and physiological traits are detailed in Lowry et al. (2015). A linkage map was generated by genotyping 247 F_2 individuals. The map contains 885 codominant markers across 1,045 cM with an intermarker map distance of ~0.5 cM. Details of linkage map construction can be found in Lovell et al. (*in review*).

Clonal replicates of the parental genotypes, their F_1 hybrid, and individual F_2 progeny were planted in the field in October 2012. The field experimental site was located in a prairie field (30.182° N, 97.879° W) at the south end of the Ladybird Johnson Wildflower Center (Austin, TX). Prior to planting, the field was covered with weed

barrier cloth (Sunbelt 3.2oz., Dewitt, Sikeston, MO, USA). For planting, holes were cut in the cloth and a mechanical auger was used to drill holes in the soil. Plants were arrayed into rows with 1.2m spacing between rows and 40cm spacing between plants, along rows. Due to exceptionally low rainfall in the fall of 2012, plants were watered as needed through November and early December to ensure establishment. Irrigation was ceased once plants entered winter dormancy. The experimental plants emerged from dormancy in the spring of 2013.

NIRS phenotyping and model building

Plant tissue for NIRS analysis was harvested at the end of the growing season in 2013. Approximately 20-40 R3 stage tillers (Hardin et al., 2013) were harvested from each of 262 F_2, 13 HAL2, 25 FIL2, and 25 F_1 plants. Tillers were dried to < 4% moisture at 50°C and knife-milled to ≤ 2-mm particle size (Thomas Model 4 Wiley Mill, Thomas Scientific, Swedesboro, NJ, USA). Samples were homogenized by riffling (Gilson Spinning Riffler SP-230, Gilson Company Inc., Lewis Center, OH, USA) for uniform particle size distribution. A Thermo Antaris II FT-NIR machine (Thermo Scientific Inc., Madison, WI, USA) was used to scan each sample following the protocol by established Sluiter et al. (2010). A subset of 50 samples was chosen from the full NIRS set for model calibration based on their distribution in the multi-spectral space. The bench-top LAPs for calibration samples are detailed in Sluiter et al. (2010) but outlined here for convenience. We performed a multi-step analytical procedure involving two-stage solvent extraction of the biomass samples (water then ethanol) followed by two-stage acid hydrolysis of the extracted biomass. We report the following values (all with units of % dry matter): water extractives, ethanol extractives, structural carbohydrates (glucan, xylan, galactan, arabinan), lignin (both acid soluble and acid-insoluble), and structural and non-structural inorganics, collectively termed “ash”. We also measured the sucrose, free glucose and free fructose concentration in the water extractives fraction. Structural

carbohydrates were measured as soluble monomeric sugars using HPLC, and then converted to a structural (anhydro) basis.

We developed a preliminary partial-least-square (PLS-2) multivariate calibration model using near-infrared spectral data for the prediction of the most abundant cell wall components: glucan, xylan, lignin, and ash. The model was fully cross-validated using the "leave-one-out" method where a single sample was removed from the model, and the model rebuilt without the sample. Model uncertainty was approximated using the root-mean-square-error of the calibration (RMSEC) and the root-mean-square-error of the cross-validation measurement (RMSECV). We then used the model to calculate a prediction for structural carbohydrates, lignin, and ash content for all remaining samples. For each sample prediction we also calculated deviation from the mean predicted value. Any sample where 2*deviation ≥ RMSEC was not used in the subsequent analysis. All LAPs, NIRS, and model building was performed at the National Renewable Energy Laboratory in Golden, CO.

QTL analysis

We mapped QTL for cell wall traits in the mapping population using a stepwise multiple-QTL model fitting method as implemented in R/qtl (Broman et al., 2003). All QTL scans were performed using a normal model and Haley-Knott regression based on a dense 0.5 cM grid of pseudomarkers generated using the calc.genoprob function. We calculated logarithmic odds-ratio (LOD) penalties for main effects and interactions for each trait through 1000 permutation of the scantwo function at an alpha of 0.1. We chose to relax the QTL significance threshold because this is an exploratory study in a new plant system and minimizing the number of false negative results is of greater importance than detecting false positive results. We conducted a forward/backward stepwise search for models with a maximum of 10 QTLs that optimized the penalized LOD score criterion. We calculated the 1.5 LOD drop interval of the QTLs in the best-fit models. We also used the best-fit stepwise model for each trait to calculate the additive effect,

dominance deviation, and percent of variance explained (PVE) for each QTL using the makeqtl and fitqtl functions of R/qtl. We calculated the phenotypic difference between parental HAL2 and FIL2 lines and the percent of parental divergence (PPD) explained by the additive effect of each QTL for the traits with significant mean differences between parents. All computation was performed on the Lonestar cluster at the Texas Advanced Computing Center at UT Austin.

Results

Cell wall compositional analysis of calibration samples

We report a full compositional analysis of the cell wall components for the 50 calibration samples in Table 3.1. As expected, the *P. hallii* cell wall is composed primarily of lignin, glucan and xylan, which together make up on average 61% of the total dry biomass. Hemicellulose is comprised of a diverse group of polysaccharides that varies in composition and covalent and non-covalent interactions with cellulose and lignin. The primary sugars that comprised hemicellulose in *P. hallii* are xylan, galactan and arabinan. Hemicellulose, including xylan, made up one fourth of the total biomass. These complicated polysaccharides are often modified by *O*-acetyl groups (Pawar et al., 2013). Acetylation can create a steric hindrance for hydrolysis and liberated acetate can inhibit fermentation by specific microorganisms (X. Chen et al., 2012). Acetyl groups were also found in lignin, to a lesser extent (Del Río et al., 2007). We report 7.34% acetyl content in the calibration samples. Water soluble sugars sucrose, glucose, and fructose were found in trace amounts and comprise 3.31% of the total biomass. Ash includes all structural and non-structural inorganic compounds such as silica, potassium, calcium, sulfur, and chlorine. Ash content can interfere with thermochemical pretreatment and negatively effect the energy density of the biomass (Laser & Lynd, 2014). We found ash to total 7.34% of the biomass. Additional extractives from the laboratory characterization

process included trace amounts of structural and non-structural polysaccharides and polymers.

P. hallii has comparable cell wall composition to other perennial grass feedstocks. We evaluated the phenotypic similarity of *P. hallii* cell wall composition traits to three C4 grasses (*Miscanthus*: Haffner et al., 2013; switchgrass: Vogel et al., 2011; *Sorghum*: Wolfrum et al., 2013) and one hard wood (poplar: Maranan & Laborie, 2008). We report the mean and standard deviation of percent dry biomass for ash, lignin, glucan, and xylan from NIRS calibration datasets with LAPs analogous to this study (Figure 3.2). The composition of *P. hallii* cell walls was comparable to switchgrass and S*orghum* across all four traits. Interestingly, *Miscanthus* fad a cell wall composition more similar to the hardwood poplar than to the other C4 grasses. We also found that relative xylan content is conserved across species (Figure 3.2).

Parent Means and Phenotypic Correlations for NIRS predicted traits

We found minimal parental divergence and significant phenotypic correlations for NIRS predicted cell wall traits. Model uncertainty was low, as indicated by RMSEC values for each trait (ash: 0.47, lignin: 0.46, glucan: 0.76, xylan: 0.67). All trait values are presented in units of percent total dry biomass. To assess parental divergence we performed a t-test comparing HAL2 and FIL2 mean trait predictions. Two of the four major cell wall components, ash and xylan, differed significantly between the parental lines (Table 3.2, Figure 3.3). We tested for heterosis by comparing parent and F_1 hybrid values. We found that the mean trait predictions for the F_1 hybrid were significantly lower than both HAL2 and FIL2 for all traits except xylan (p-value < 0.05). We found some evidence for transgressive segregation in two traits without significant parental divergence: glucan and lignin. For these traits, the range of the F_2 population exceeded that of the mean predictions for HAL2 and FIL2 (Figure 3.3).

To explore patterns of trait correlations we calculated phenotypic correlations among traits for parent lines and for the recombinant F_2 progeny. We found the

significance of the correlations did not change across generations so we only report the correlations for the mapping population. The three major cell wall structural components, glucan, xylan, and lignin, were positively correlated. Ash was negatively correlated with lignin, and xylan and has no significant correlation with glucan (Table 3.3).

QTL analysis

We identified 14 QTL and one epistatic interaction from 4 NIRS-predicted cell wall traits using stepwise model selection (Figure 3.4, Table 3.4). We found the most QTL for ash (7) and detected at least one QTL for all the cell wall traits at an alpha of 0.1. We found 6 QTL that colocalized to linkage group (LG) 3. We also found colocalized QTL for pairs of traits on LGs 5, 8 and 9. Ash and xylan had a significant negative phenotypic correlation and collocated QTL on both LGs 5 and 8. The allelic effects for these QTL were all in the expected direction of divergence, where the HAL2 allele increased ash content and decreased xylan content, at each location. We found only two occurrences where the allelic effects were not in the expected direction of parental divergence. These occur on LG 1 and LG 3 for ash. The QTL for ash on LG 3 had an epistatic interaction and contrasting allelic effects with substantial dominance.

The largest percent variance explained (PVE) by a single additive QTL was 10% for ash on LG 3. However, two additive ash QTL, which also shared an epistatic interaction, explained 18.6% and 26% of the variance. Additive effects ranged from 0.06-0.9 percent total biomass and dominance effects ranged from 0.03-0.5 percent total biomass. We calculated the mean difference between HAL2 and FIL2 for ash and xylan and found that the percent of parental divergence (PPD) explained by each locus ranged from 9.33-64.79% for ash and from 42.82-53.52% for xylan.

Discussion

We found almost half of cell wall QTL localized to LG 3 in *P. hallii*. In addition, we discovered significant positive correlations between cellulose, hemicellulose, and lignin suggesting a potential for pleiotropic interactions among these traits. These results have implications for recalcitrance and the trade-off of diverting energy to sugar storage or building a structural matrix. Production of cellulose, hemicellulose, and lignin is sequential in the life stages of a plant. Cellulose is the first structure to form in new cell walls, then hemicellulose is added to the cellulose with a series of covalent and non-covalent bonds, and lignin is the last component to be made across cell wall development. The hardening of a plant stem late in life is called lignification. Understanding the timing and trade-offs in this process is important for biomass harvest so as to avoid unnecessary hindrance to fuel conversion. Interestingly, using the same *P. hallii* F_2 mapping population, Lowry et al. (2015) localized 9 QTL to LG 3. These QTL included morphological traits such as tiller, leaf, and reproductive characteristics, and physiological traits such as CO_2 assimilation rate and stomatal conductance. The density of functional alleles in this chromosome region suggests the possibility of a major developmental regulator controlling growth with concomitant impacts on important cell wall characteristics.

It is not surprising that we found the largest and most numerous QTL in our study associated with ash content. Ash content is a composite trait driven by the accumulation of inorganic molecules in plant tissues (Monti et al., 2008). Generally, a large ash content reduces the overall density of energy convertible material in biomass. However, ash is inclusive of all structural and non-structural inorganic material and the type and function of inorganic molecules in plants are diverse. Wang et al. (2014) found that switchgrass ash was composed of 67% SiO_2, 12% CaO and many other mineral oxides in small concentrations. High silica content can melt and fuse together when biomass is thermochemically pretreated before enzyme digestion (Laser & Lynd, 2014), thereby causing problems when scaling up fuel conversion methods at a biorefinery. However, if harvested, the inorganic residue can be recycled into fertilizer (Schiemenz & Eichler-Löbermann, 2010) or a cement additive (Rajamma et al., 2009; Wang et al., 2014). Silica

content has many functions in plant tissue including defense from abiotic and biotic stress (Sangster & Hodson, 1986; Tripathi et al., 2014). While there is mixed evidence regarding trade-offs in silica and carbon-based defenses (Cooke & Leishman, 2012; Moles et al., 2013), silica has been found to reduce herbivory in both switchgrass and *Miscanthus* (Nabity et al., 2012).

Hemicellulose is a complex polymer made up of several different 5- and 6-carbon polysaccharides with varying amounts of methylation, acetylation, and covalent bonds (Jeffries, 1994). This is in stark contrast to the uniform cellulose polymer composed entirely of glucan monomers. Xylan is the primary structural pentose in monocotyledon hemicellulose and we found that the parental HAL2 line contained, on average, 1.1 mg/g less xylan than FIL2. We found no significant parental divergence for either glucan or lignin. We also observed that *filipes* stems are more upright and erect in stature than *hallii* (Figure 3.1). Hemicellulose is one of only three main structural components in the cell wall therefore removal of any component would presumably lead to less structural integrity of stem tissue. However, there is tantalizing evidence that xylan is linked to xylem production. Brown et al. (2007) found that *Arabadopsis* xylan knockouts result in dwarfing and collapsed xylem. Reduced xylan in rice and corn leads to a droopy stature (S.-J. Zhang et al., 2012). This suggests there may be a link between the xylan content and vascular tissue in *P. hallii.*

This is only the second study to use NIRS for QTL mapping of cell wall traits. Genetic mapping requires large sample sizes and therefore phenotyping needs to be fast, accurate, and inexpensive. NIRS is used to predict plant tissue composition for two types of applications in biofuels research. It is used to quantify raw cell wall components from untreated biomass, such as in this study, and it is used to analyze different steps in the biofuel conversion process such as digestibility (Hou & Li, 2011; Huang et al., 2012; Roberts et al., 2011) and ethanol yield (Guimarães et al., 2014; Lorenz, Anex, et al., 2009). We argue that while this second category is a valuable avenue of research, metrics such as *in vitro* dry matter digestibility (IVDMD), glucose release (GLCRel), and ethanol yield confound the efficiency of conversion method with feedstock characteristics. It is

important to have a clear understanding of the initial tissue composition and how it functions within the plant as conversion methods are still in development.

Lignocellulosic feedstock species vary in cell wall quality and composition as well as in availability of genomic resources for crop improvement. Many of the well-developed grass genome resources come from long domesticated crops such as corn, sugarcane, and sorghum. These crops have been bred for high-quality land use and intensive agricultural practices. Whereas perennial grasses such as switchgrass, *Miscanthus* and *Andropogon* are only recently being cultivated for agriculture and their genomes have not been subject to 8-10,000 years of anthropogenic selection. Development of *P. hallii* as a model system has contributed to assembly of the switchgrass genome. We compared *P. hallii* to three other energy grasses all within the same subfamily (Panicoideae) and one hardwood species. We expected the grasses to have similar cell wall composition and found that *P. hallii* did closely resemble switchgrass and *Sorghum*. Strikingly, *Miscanthus* seems to have a cell wall composition that more closely resembles the hardwood, poplar. Cell wall composition of hardwoods and grasses differs in primary and secondary cell wall characteristics. Hardwoods produce more primary cell wall structure, which contains pectin, another polysaccharide component, in addition to cellulose and hemicellulose. Grasses have substantially more secondary cell wall structure, which has a higher lignin content (DeMartini et al., 2013). Variation in types and composition of lignocellulosic feedstocks is advantageous in that is broadens the range of acceptable habitats that can be used to produce substantial biomass material to meet global demand for renewable energy sources. However, diversity in the composition and quality of lignocellulosic feedstocks is disadvantageous for the biofuel conversion process because biorefineries require uniform and high quality biomass. Understanding the underlying genetic architecture of these quality biomass composition traits will lead to a better understanding of the structure and function of cell walls and contribute to future breeding efforts and crop development of bioenergy feedstocks.

Tables

Composition Component	Mean	SE	Max.	Min.
% Total Ash	7.34	0.23	11.1	4.2
% Structural Inorganics	3.75	0.17	6.7	1.4
% Non-Structural Inorganics	3.6	0.09	5.1	2.1
% Total Protein	1.23	0.03	1.7	0.9
% Sucrose	1.57	0.21	6.3	0
% Free Glucose	0.74	0.07	2.2	0.1
% Free Fructose	1.0	0.11	5.2	0.1
% Water Extractable Others	10.27	0.26	13.7	6.3
% Ethanol Extractives	3.83	0.06	4.8	3.2
% Lignin	14.37	0.13	15.7	12.2
% Glucan	28.8	0.19	32.8	25.6
% Xylan	18.54	0.14	21.1	16.7
% Galactan	1.73	0.06	3.0	1.2
% Arabinan	3.74	0.07	5.4	2.6
% Acetyl	7.34	0.23	11.1	4.2
Total %	93.9	0.2	96.9	91.2

Table 3.1 Cell wall composition statistics

50 calibration samples reported as % of total dry biomass. SE- standard error, Max- maximum value, Min- minimum value.

	Ash	Lignin	Glucan	Xylan	N
F2	7.01 (0.08)	14.29 (0.04)	28.37 (0.07)	18.5 (0.04)	262
F2 range	3-11.43	12.68-16.55	25.26-31.77	17-20.88	--
F1	6.7 (0.11)	14.3 (0.09)	28.28 (0.17)	18.58 (0.09)	25
FIL2	7.26 (0.13)	14.95 (0.13)	29.03 (0.21)	18.77 (0.11)	25
HAL2	10.13 (0.25)	14.77 (0.16)	29.3 (0.23)	17.67 (0.13)	13
pval	<0.0001	0.405	0.411	<0.0001	--

Table 3.2 Cell wall trait predictions

Cell wall trait predictions, mean (SE). N is the number of replicates for the parental lines and the number of F_2 individuals measured for each trait, pval significance is the result of a t-test for difference between the parental (HAL2 and FIL2) lines. Trait values are % dry biomass.

Trait 1	Trait 2	r	pval
Ash	Lignin	-0.23	<0.0001
Ash	Glucan	0.07	0.24
Ash	Xylan	-0.54	<0.0001
Lignin	Glucan	0.54	<0.0001
Lignin	Xylan	0.53	<0.0001
Glucan	Xylan	0.50	<0.0001

Table 3.3 Phenotypic trait correlations for *P*. *hallii* F_2 population

Pairwise Pearson product-moment correlation (r) and significance (pval).

Trait	LG	Pos	1.5- LOD	LOD	a (SE)	D (SE)	PVE	PPD
Ash	1	61.17	56-65.2	7.73	-0.449 (0.076)	-0.026 (0.102)	5.55	-31.23
Ash *	3	5.08	3.5-6.9	28.88	0.847 (0.091)	-0.198 (0.119)	26.06	58.87
Ash *	3	71.11	69.7-73.8	22.17	-0.932 (0.092)	-0.145 (0.117)	18.57	-64.79
Ash	4	36.48	31.4-39	5.06	0.134 (0.075)	0.493 (0.105)	3.53	9.33
Ash	5	1.63	0-4.7	9.15	0.563 (0.088)	-0.136 (0.11)	6.67	39.16
Ash	8	47.65	42.1-50.7	13.09	0.629 (0.079)	0.132 (0.107)	9.95	43.74
Ash	9	161.45	131.5-167.1	6.57	0.379 (0.07)	-0.092 (0.103)	4.66	26.31
Glucan	3	61.07	37-71.1	3.64	-0.398 (0.097)	-0.13 (0.136)	7.14	NA
Lignin	3	21.15	16.3-28.1	4.48	-0.187 (0.054)	-0.246 (0.074)	7.57	NA
Lignin	3	77.27	61.1-85.4	4.38	-0.224 (0.056)	-0.18 (0.074)	7.39	NA
Lignin	9	167.14	162.3-169.2	3.55	0.064 (0.051)	-0.291 (0.074)	5.94	NA
Xylan	3	26.24	21.1-29.7	5.84	-0.293 (0.059)	0.082 (0.078)	9.62	53.52
Xylan	5	6.52	0-20.4	4.80	-0.275 (0.059)	0.049 (0.082)	7.82	50.26
Xylan	8	56.39	46-66.4	4.77	-0.234 (0.058)	-0.187 (0.077)	7.76	42.82

Table 3.4 QTL and main effects for cell wall traits in *P. hallii*

LG- linkage group, Pos- position of QTL on LG in cM, 1.5-LOD- 1.5 LOD drop confidence interval for each QTL in cM, LOD- logarithm of odds score, a (SE)-additive effect and standard error, D (SE)- dominance deviation and standard error, PVE- percent of additive variance explained by each QTL, PPD- percent of parental divergence explained if applicable.
*epistatic interaction

Figures

Figure 3.1 *P. hallii* varieties

P. hallii var. *filipes* (left) and *P. hallii* var. *hallii* (right).

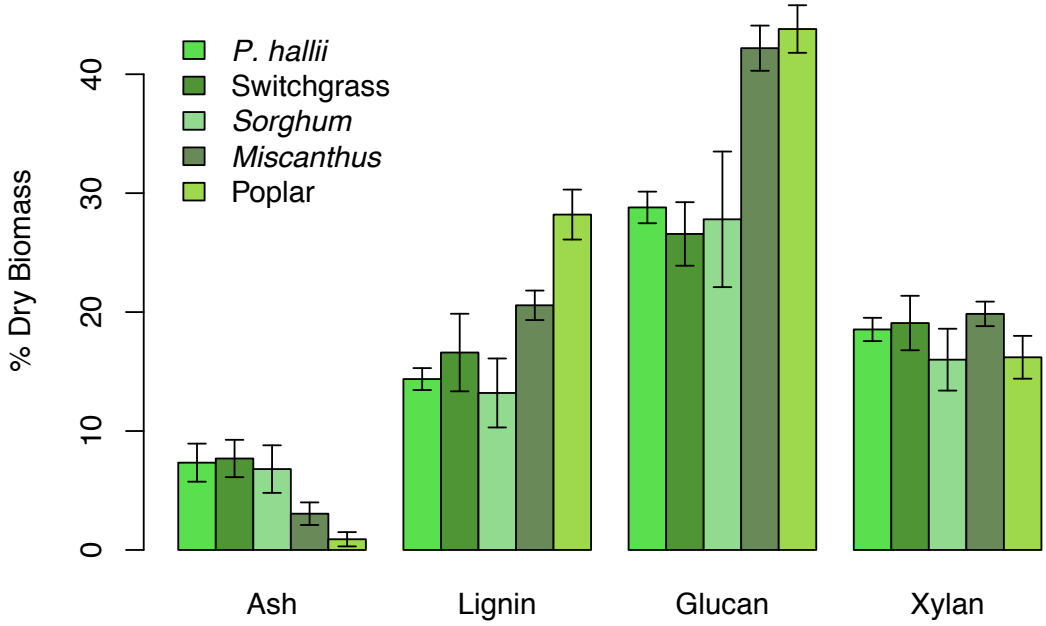

Figure 3.2 Cell wall composition for lignocellulosic feedstocks

Comparison of major cell wall composition components in five lignocellulosic plant species. Mean and standard deviation values for NIRS calibration datasets are reported for *P. hallii*, switchgrass (Vogel et al., 2011), *Sorghum* (Wolfrum et al., 2013), *Miscanthus* (Haffner et al., 2013), and poplar (Maranan & Laborie, 2008).

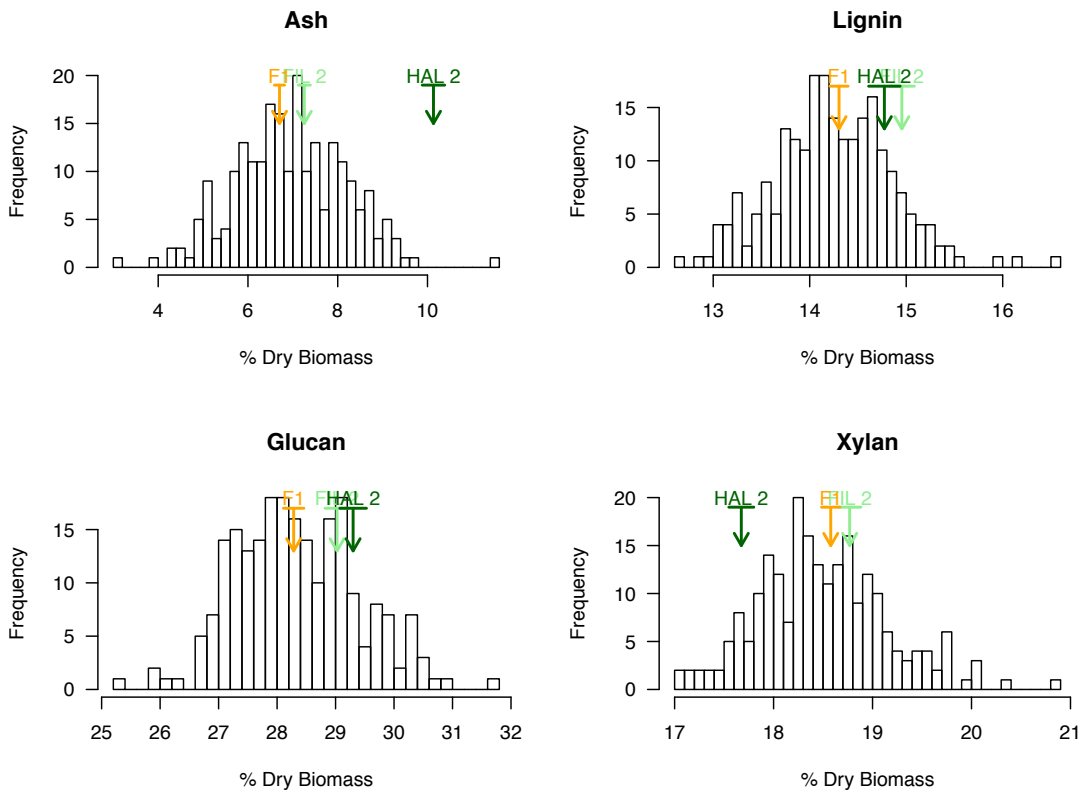

Figure 3.3 Phenotypic trait distributions for *P. hallii* F_2 mapping population

Parent and F_1 hybrid means are indicated by vertical arrows and standard error indicated by horizontal line.

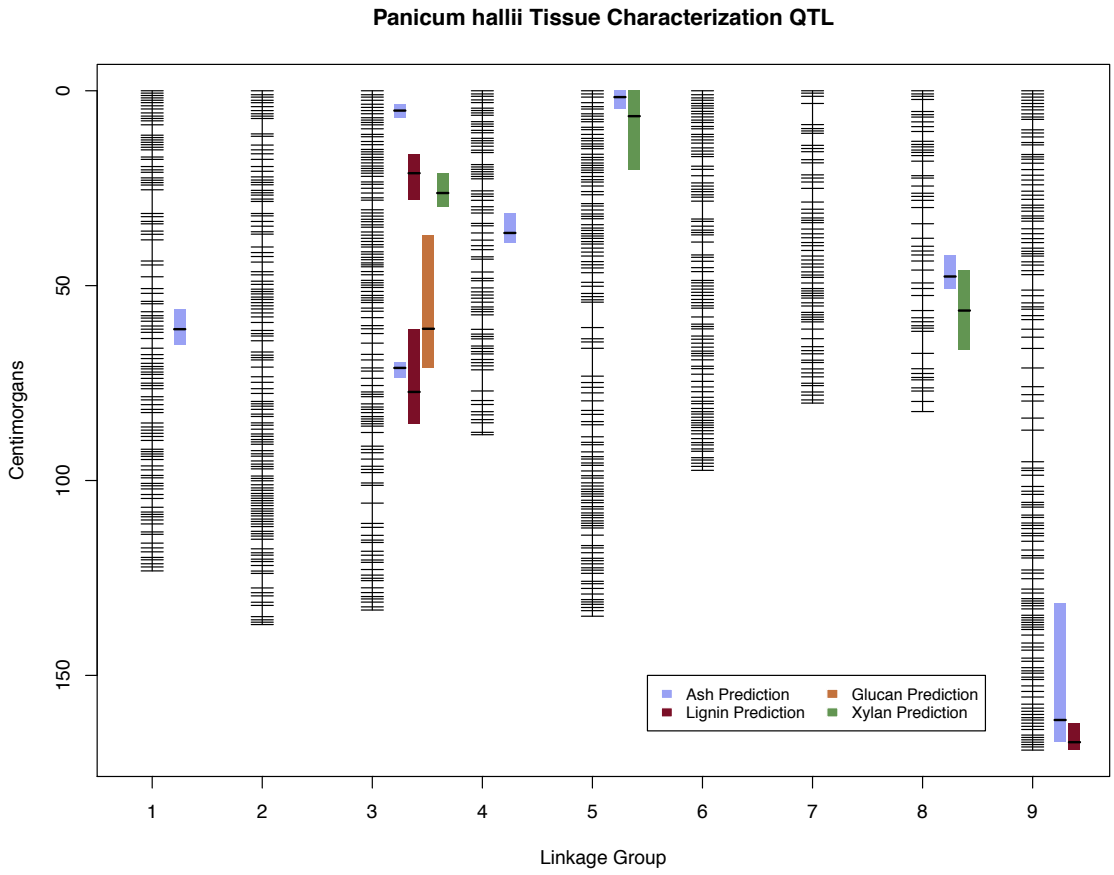

Figure 3.4 *P. hallii* genetic linkage map with cell wall QTL

QTL plotted to the left of respective linkage groups. Color bars represent 1.5-LOD interval and horizontal line indicates location of QTL.

Appendix

Supplemental material for Chapter 1.

Collection Site	Sample size	Latitude (N)	Longitude (W)	Analysis Dataset	Population ID
Mtn. Research St.	9	40.05	-105.57	Microsatellite Markers	1
Lefthand Canyon	7	40.02	-105.5	Microsatellite Markers	2
Gross Dam Rd.	10	39.95	-105.25	Microsatellite Markers	3
Rt. 72	10	39.94	-105.39	Microsatellite Markers	4
Copperdale Inn	10	39.92	-105.09	Microsatellite Markers	5
Rollinsville Rd.	10	39.92	-105.35	Microsatellite Markers	6
Blackhawk	10	39.78	-105.35	Microsatellite Markers	7
Bailey	10	39.41	-105.42	Microsatellite Markers	8
Spring Creek	10	39.35	-105.08	Microsatellite Markers	9
Wilkerson Pass	14	39.05	-105.55	Microsatellite Markers	10
Cuchara	13	37.33	-105.38	Microsatellite Markers	11
E. Bailey	10	39.55	-105.52	Microsatellite Markers	M1
Elbert	10	37.55	-105.48	Microsatellite Markers	M2
Foxton	10	37.73	-105.5	Microsatellite Markers	M3
La Veta	11	37.25	-105.42	Microsatellite Markers	M4
Mosca	11	39.45	-105.5	Microsatellite Markers	M5
North Fork	14	37.09	-105.55	Microsatellite Markers	M6
Pine Valley Rd.	11	39.48	-105.69	Microsatellite Markers	M7
San Francisco	10	37.61	-105.45	Microsatellite Markers	M8
Scaffer's Crossing	10	38.74	-105.25	Microsatellite Markers	M9
N. La Veta Pass	10	39.19	-105.46	Microsatellite Markers	M10
West Nathrop	10	39.44	-105.52	Microsatellite Markers	M11
Mtn. Research St.	20	40.03	-105.39	Quantitative Traits	1
Lefthand Canyon	11	40.07	-105.45	Quantitative Traits	2
Gross Dam Rd.	7	39.94	-105.42	Quantitative Traits	3
Rt. 72	22	39.94	-105.52	Quantitative Traits	4
Copperdale Inn	16	39.9	-105.48	Quantitative Traits	5
Rollinsville Rd.	12	39.98	-105.09	Quantitative Traits	6*
Caribou Rd.	6	39.97	-105.42	Quantitative Traits	6*
Blackhawk	19	39.81	-105.08	Quantitative Traits	7
Bailey	3	39.48	-105.69	Quantitative Traits	8
Spring Creek	9	39.36	-105.5	Quantitative Traits	9
Wilkerson Pass	14	39.05	-105.25	Quantitative Traits	10
Cuchara	10	37.33	-105.5	Quantitative Traits	11
Noe	6	39.18	-105.35	Quantitative Traits	Q1
Hwy 86	11	39.38	-105.38	Quantitative Traits	Q2
West Chicago Cr.	5	39.69	-105.5	Quantitative Traits	Q3
Georgetown	6	39.76	-105.55	Quantitative Traits	Q4
Hwy105	3	39.43	-105.35	Quantitative Traits	Q5

Table A1 Collection site information

Sample size for the quantitative trait dataset indicates the number of half-sib families, sample size for the microsatellite dataset indicates number of individuals. Populations with matching population IDs indicate the subset of populations used for the driftsel analysis. *sites were merged into a single site for use in *driftsel* analysis.

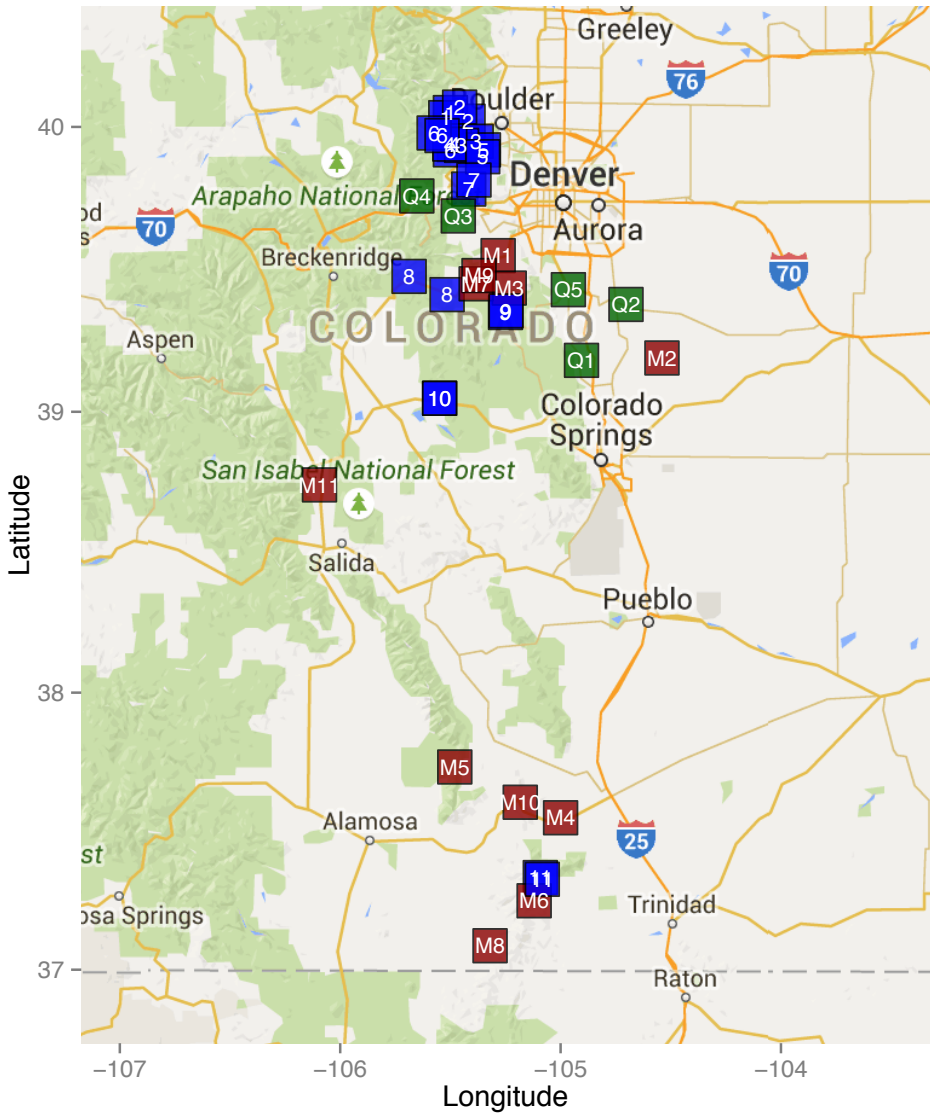

Figure A1 Map of collection sites

Map of all collection sites in the Front Range of the Colorado Rockies. Number refers to population ID in Table A1. Blue color represents 'matched' quantitative and molecular sites used in *driftsel* analysis. Green color represents sites with only quantitative data. Red represents sites with only molecular data.

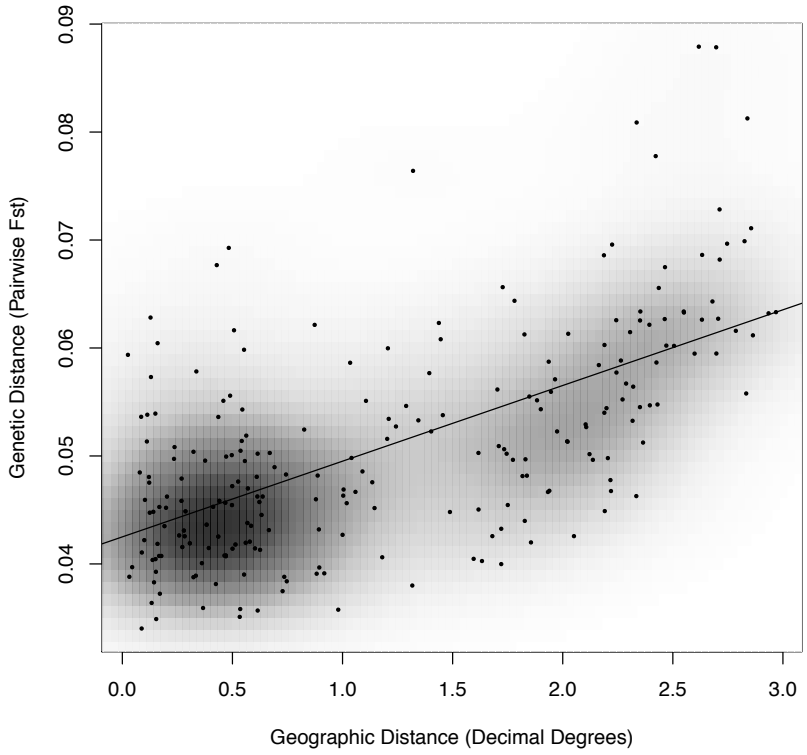

Figure A2 Nuclear isolation by distance

Isolation by distance plot with density shading for nuclear markers.

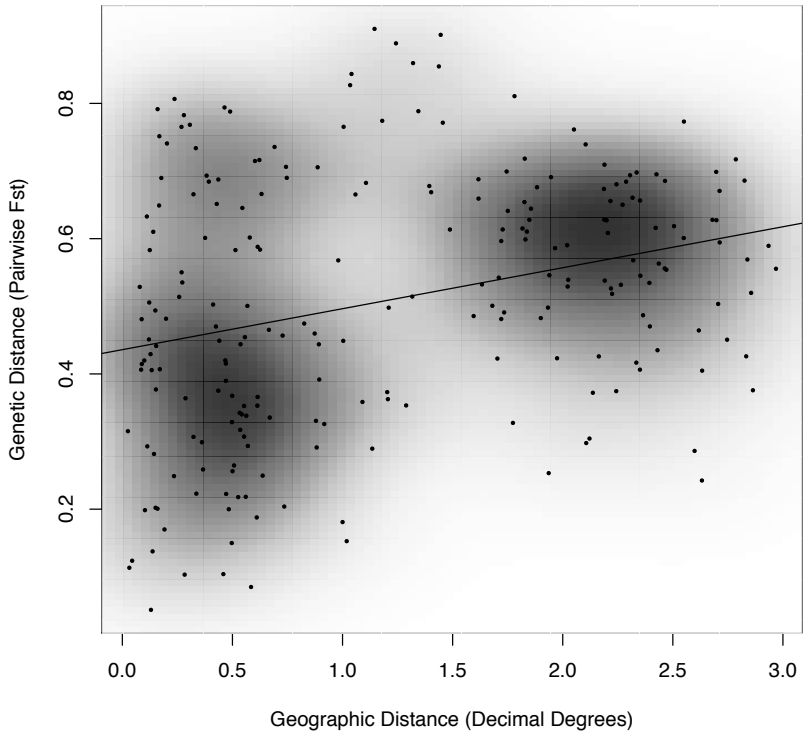

Figure A3 Chloroplast isolation by distance

Isolation by distance plot with density shading for chloroplast markers.

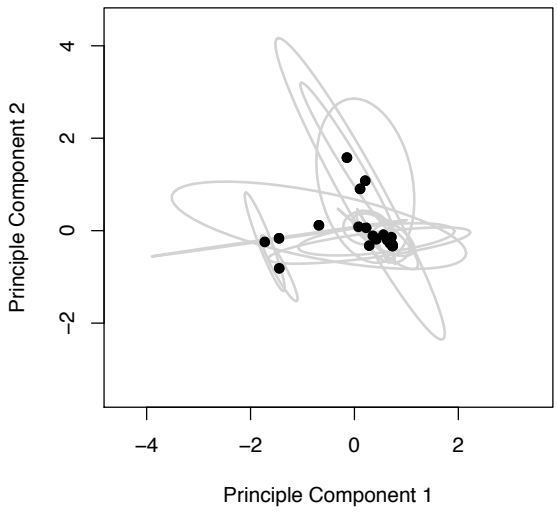

Figure A4 Chloroplast microsatellite PCA

Principal components (PCs) 1 and 2 for chloroplast microsatellite loci. Dots represent population means, ellipses indicate 95% CI.

References

Amaducci, S., Amaducci, M. T., Benati, R., & Venturi, G. (2000). Crop yield and quality parameters of four annual fibre crops (hemp, kenaf, maize and sorghum) in the North of Italy. *Industrial Crops and Products*, *11*(2-3), 179–186.

Åman, P. (1993). Composition and Structure of Cell Wall Polysaccharides in Forages. In R. D. H. and J. R. H. G. Jung, D. R. Buxton (Ed.), *Forage Cell Wall Structure and Digestibility* (pp. 183–199). Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Atsumi, S., Hanai, T., & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, *451*(7174), 86–89.

Barney, J. N., Mann, J. J., Kyser, G. B., Blumwald, E., Van Deynze, A., & DiTomaso, J. M. (2009). Tolerance of switchgrass to extreme soil moisture stress: Ecological implications. *Plant Science*, *177*(6), 724–732.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*(1), 38–44.

Beavis, W. D. (1998). QTL analyses: power, precision, and accuracy. In P. A. H. (Ed.), *Molecular Dissection of Complex Traits* (pp. 145–162). New York: CRC Press.

Bergelson, J., & Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature reviews. Genetics*, *11*(12), 867–879.

Blows, M. W. (2007). A tale of two matrices: multivariate approaches in evolutionary biology. *Journal of Evolutionary Biology*, *20*, 1–8.

Blows, M. W., & Hoffmann, A. a. (2005). A reassessment of genetic limits to evolutionary change. *Ecology*, *86*(6), 1371–1384.

Bouton, J. H. (2007). Molecular breeding of switchgrass for use as a biofuel crop. *Current Opinion in Genetics and Development*, *17*(6), 553–558.

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890.

Brown, D. M., Goubet, F., Wong, V. W., Goodacre, R., Stephens, E., Dupree, P., & Turner, S. R. (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant Journal*, *52*(6), 1154–1168.

Buxton, D. R., & Casler, M. D. (1993). Environmental and Genetic Effects on Cell Wall Composition and Digestibility. In J. Jung, H. G.; Buxton, D. R.; Hatfield, R. D.; Ralph (Ed.), *Forage Cell Wall Structure and Digestibility* (pp. 685–714). Maidson, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Campbell, D. R. (1989). Measurements of selection in a hermaphroditic plant: variation in male and female pollination success. *Evolution*, *43*(2), 318–334.

Campbell, D. R. (1991). Effects of Floral Traits on Sequential Components of Fitness in Ipomopsis aggregata. *The American Naturalist*, *137*(6), 713.

Campbell, D. R. (1996). Evolution of Floral Traits in a Hermaphroditic Plant: Field Measurements of Heritabilities and Genetic Correlation. *Evolution*, *50*(4), 1442–1453.

Campbell, D. R. (1997). Genetic correlation between biomass allocation to male and female functions in a natural population of Ipomopsis aggregata. *Heredity*, *79*, 606–614.

Campbell, D. R. (2003). Natural selection in Ipomopsis hybrid zones: Implications for ecological speciation. *New Phytologist*, *161*(1), 83–90.

Campbell, D. R. (2009). Using phenotypic manipulations to study multivariate selection of floral trait associations. *Annals of botany*, *103*(9), 1557–1566.

Campbell, D. R., Waser, N. M., & Melendez-Ackerman, E. J. (1997). Analyzing pollinator-mediated selection in a plant hybrid zone: hummingbird visitation patterns on three spatial scales. *American Naturalist*, *149*(2), 295–315.

Campbell, D. R., Waser, N. M., & Price, M. V. (1996). Mechanisms of hummingbird-mediated selection for flower width in Ipomopsis aggregata. *Ecology*, *77*(5), 1463–1472.

Campbell, D. R., Waser, N. M., & Wolf, P. G. (1998). Pollen Transfer By Natural Hybrids and Parental Species in an Ipomopsis Hybrid Zone. *Evolution*, *52*(6), 1602–1611.

Carriquiry, M. a., Du, X., & Timilsina, G. R. (2011). Second generation biofuels: Economics and policies. *Energy Policy*, *39*(7), 4222–4234.

Casler, M. (2012). Switchgrass Breeding, Genetics, and Genomics. In A. Monti (Ed.), *Switchgrass* (pp. 29–54). London: Springer.

Casler, M. D., & Boe, a. R. (2003). Cultivar × Environment Interactions in Switchgrass. *Crop Science*, *43*(6), 2226.

Casler, M. D., Tobias, C. M., Kaeppler, S. M., Buell, C. R., Wang, Z.-Y., Cao, P., … Ronald, P. (2011). The switchgrass genome: tools and strategies. *The Plant Genome*, *4*(3), 273–282.

Casler, M. D., Vogel, K. P., Taliaferro, C. M., Ehlke, N. J., Berdahl, J. D., Brummer, E. C., … Mitchell, R. B. (2007). Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Science*, *47*(6), 2249–2260.

Casler, M., & Vogel, K. (2004). Latitudinal adaptation of switchgrass populations. *Crop Science*, *44*, 293–303.

Cassida, K. a., Muir, J. P., Hussey, M. a., Read, J. C., Venuto, B. C., & Ocumpaugh, W. R. (2005a). Biofuel component concentrations and yields of switchgrass in south central U.S. environments. *Crop Science*, *45*(2), 682–692.

Cassida, K. a., Muir, J. P., Hussey, M. a., Read, J. C., Venuto, B. C., & Ocumpaugh, W. R. (2005b). Biomass yield and stand characteristics of switchgrass in south central U.S. environments. *Crop Science*, *45*(2), 673–681.

Catchen, J. M., Amores, A. A., Hohenlohe, P. A., Cresko, W. W., & Postlethwait, J. H. J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182.

Chen, F., & Dixon, R. a. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nature biotechnology*, *25*(7), 759–761.

Chen, X., Shekiro, J., Franden, M. A., Wang, W., Zhang, M., Kuhn, E., … Tucker, M. P. (2012). The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process. *Biotechnology for Biofuels*, *5*(1), 8.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal Jr, G., Dickson, M., Grimwood, J., … Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*(5717), 1928.

Cooke, J., & Leishman, M. R. (2012). Tradeoffs between foliar silicon and carbon-based defences: Evidence from vegetation communities of contrasting soil types. *Oikos*, *121*(12), 2052–2060.

Cortese, L. M., Honig, J., Miller, C., & Bonos, S. a. (2010). Genetic diversity of twelve switchgrass populations using molecular and morphological markers. *Bioenergy Research*, *3*(3), 262–271.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. *Sunderland, MA: Sinauer Associates*. Sunderland, MA: Sinauer Associates.

Del Río, J. C., Marques, G., Rencoret, J., Martínez, Á. T., & Gutiérrez, A. (2007). Occurrence of naturally acetylated lignin units. *Journal of Agricultural and Food Chemistry*, *55*(14), 5461–5468.

DeMartini, J. D., Pattathil, S., Miller, J. S., Li, H., Hahn, M. G., & Wyman, C. E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy & Environmental Science*, *6*(3), 898–909.

Des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (2013). Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 5–29.

Dien, B. S., Cotta, M. a., & Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: Current status. *Applied Microbiology and Biotechnology*, *63*(3), 258–266.

Dien, B. S., Jung, H. J. G., Vogel, K. P., Casler, M. D., Lamb, J. F. S., Iten, L., … Sarath, G. (2006). Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass and Bioenergy*, *30*(10), 880–891.

Earl, D. a., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361.

Eberhart, S. a., & Newell, L. C. (1959). Variation in Domestic Collections of Switchgrass, Panicum virgatum L1. *Agronomy Journal*, *51*(10), 613.

Elam, D., & Linhart, Y. (1988). Pollination and Seed Production in Ipomopsis aggregata: Differences Among and Within Flower Color Morphs. *American Journal of Botany*, *75*(9), 1262–1274.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Faegri, K., & Pijl, L. van der. (1979). *The principles of pollination ecology, Volume 1978*. Elsevier Science Limited.

Fenster, C., Armbruster, W., Wilson, P., Dudash, M., & Thomson, J. (2004). Pollination Syndromes and Floral Specialization. *Annual Review of EcologyEvolutionand Systematics*, *35*, 375–403.

Forkmann, G. (1991). Flavonoids as Flower Pigments: The Formation of the Natural Spectrum and its Extension by Genetic Engineering. *Plant Breeding*, *106*(1), 1–26.

Gilbert, K. J., & Whitlock, M. C. (2014). QST-FST comparisons with unbalanced half-sib designs. *Molecular Ecology Resources*, *15*, 262–267.

Glennon, K. L., Rissler, L. J., & Church, S. a. (2012). Ecogeographic isolation: A reproductive barrier between species and between cytotypes in Houstonia (Rubiaceae). *Evolutionary Ecology*, *26*(4), 909–926.

Gow, J. L., Peichel, C. L., & Taylor, E. B. (2006). Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Molecular Ecology*, *15*(3), 739–752.

Grant, V. (1949). Pollination systems as isolating mechanisms in angiosperms. *Evolution*, *3*(1), 82–97.

Grant, V. (1956). A synopsis of Ipomopsis. *Aliso*, *3*, 351–362.

Grant, V. (1981). *Plant speciation, 2nd edition. New York, New York, USA*. NewYork, New York: Columbia University Press.

Grant, V. (1992a). Floral isolation between ornithophilous and sphingophilous species of Ipomopsis and Aquilegia. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(24), 11828.

Grant, V. (1992b). Systematics and Phylogeny of the Ipomopsis aggregata Group (Polemoniaceae): Traditional and Molecular Approaches. *Systematic Botany*, *17*(4), 683–691.

Grant, V. (1994). Modes and origins of mechanical and ethological isolation in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(1), 3–10.

Grant, V., & Grant, K. A. (1965). *Flower pollination in the phlox family*. New York: Columbia University Press.

Grant, V., & Wilken, D. (1987). Secondary Intergradation between Ipomopsis aggregata candida and I. a. collina (Polemoniaceae) in Colorado. *Botanical Gazette*, *148*(3), 372–378.

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Plant Biology*, *57*, 761–780.

Guimarães, C. C., Simeone, M. L. F., Parrella, R. a C., & Sena, M. M. (2014). Use of NIRS to predict composition and bioethanol yield from cell wall structural components of sweet sorghum biomass. *Microchemical Journal*, *117*, 194–201.

Haffner, F. B., Mitchell, V. D., Arundale, R. a., & Bauer, S. (2013). Compositional analysis of Miscanthus giganteus by near infrared spectroscopy. *Cellulose*, *20*(4), 1629–1637.

Haley, C. S., Knott, S. a., & Elsen, J. M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics*, *136*(3), 1195–1207.

Harborne, J. B., & Smith, D. M. (1978). Correlations between anthocyanin chemistry and pollination ecology in the polemoniaceae. *Biochemical Systematics and Ecology*, *6*(2), 127–130.

Hardin, C. F., Fu, C., Hisano, H., Xiao, X., Shen, H., Stewart, C. N., … Wang, Z. Y. (2013). Standardization of Switchgrass Sample Collection for Cell Wall and Biomass Trait Analysis. *Bioenergy Research*, *6*(2), 755–762.

Harvell, C., Mitchell, C., Ward, J., Altizer, S., Dobson, A., Ostfeld, R., & Samuel, M. (2002). Climate warming and disease risks for terrestrial and marine biota. *Science*, *296*(June), 2158–2163.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American naturalist*, *173*(5), 579–588.

Himmel, M. E., Ding, S., Johnson, D. K., Adney, W. S., Nimlos, M., Brady, J., & Foust, T. (2007). Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, *315*(February), 804–807.

Hoeksema, J. D., & Forde, S. (2008). A meta-analysis of factors affecting local adaptation between interacting species. *American Naturalist*, *171*(3), 275–290.

Hopkins, A., Vogel, K. P., Moore, K. J., Johnson, K. D., & Carlson, I. T. (1995a). Genotype effects and genotype by environment interactions for traits of elite switchgrass populations. *Crop Science*, *35*(1), 125–132.

Hopkins, A., Vogel, K. P., Moore, K. J., Johnson, K. D., & Carlson, I. T. (1995b). Genotypic variability and genotype X environment interactions among switchgrass accessions from the midwestern USA. *Crop Science*, *35*(2), 565–571.

Hopkins, R., Levin, D. a, & Rausher, M. D. (2011). Molecular Signatures of Selection on Reproductive Character Displacement of Flower Color in Phlox Drummondii. *Evolution*, *66*, 469–485.

Hou, S., & Li, L. (2011). Rapid Characterization of Woody Biomass Digestibility and Chemical Composition Using Near-infrared Spectroscopy. *Journal of Integrative Plant Biology*, *53*(2), 166–175.

Huang, J., Xia, T., Li, A., Yu, B., Li, Q., Tu, Y., … Peng, L. (2012). A rapid and consistent near infrared spectroscopic assay for biomass enzymatic digestibility upon various physical and chemical pretreatments in Miscanthus. *Bioresource Technology*, *121*, 274–281.

Jeffries, T. W. (1994). Biodegradation of lignin and hemicelluloses. In C. Ratledge (Ed.), *Biochemistry of microbial degradation* (pp. 233–277). Springer Netherlands.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, *24*(11), 1403–1405.

Jombart, T., & Devillard, S. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(94).

Juenger, T., & Bergelson, J. (1998). Pairwise versus diffuse natural selection and the multiple herbivores of scarlet gilia, Ipomopsis aggregata. *Evolution*, *52*(6), 1583–1592.

Jung, H. G., & Allen, M. S. (1995). Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. *Journal of animal science*, *73*(9), 2774–2790.

Karhunen, M., Merilä, J., Leinonen, T., Cano, J. M., & Ovaskainen, O. (2013). driftsel: An R package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources*, *13*(4), 746–754.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241.

Kay, K. M. (2006). Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. *Evolution; international journal of organic evolution*, *60*(3), 538–552.

Kirkpatrick, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, *136*(2), 271–284.

Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical research*, *26*(5-6), 221–235.

Lande, R. (1992). Neutral Theory of Quantitative Genetic Variance in an Island Model with Local Extinction and Colonization. *Evolution*, *46*(2), 381–389.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*(6), 1210–1226.

Laser, M., & Lynd, L. (2014). Introduction to Cellulosic Energy Crops. In D. L. Karlen (Ed.), *Cellulosic Energy Cropping Systems*. West Sussex: John Wiley & Sons.

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*, 1548–1566.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*(12), 1–8.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*(4), 183–189.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*, *00*(00), 3.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595.

Lorenz, A. J., Anex, R. P., Isci, A., Coors, J. G., de Leon, N., & Weimer, P. J. (2009). Forage quality and composition measurements as predictors of ethanol yield from maize (Zea mays L.) stover. *Biotechnology for biofuels*, *2*, 5.

Lorenz, A. J., Coors, J. G., De Leon, N., Wolfrum, E. J., Hames, B. R., Sluiter, D., & Weimer, P. J. (2009). Characterization, genetic variation, and combining ability of maize traits relevant to the production of cellulosic ethanol. *Crop Science*, *49*(1), 85–98.

Lorenzana, R. E., Lewis, M. F., Jung, H. J. G., & Bernardo, R. (2010). Quantitative trait loci and trait correlations for maize stover cell wall composition and glucose release for cellulosic ethanol. *Crop Science*, *50*(2), 541–555.

Lowry, D. B. (2012). Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society*, *106*(2), 241–257.

Lowry, D. B., Behrman, K. D., Grabowski, P., Morris, G. P., Kiniry, J. R., & Juenger, T. E. (2014). Adaptations between ecotypes and along environmental gradients in Panicum virgatum. *The American naturalist*, *183*(5), 682–92.

Lowry, D. B., Hernandez, K., Taylor, S. H., Meyer, E., Logan, T. L., Barry, K. W., … Juenger, T. E. (2015). The genetics of divergence and reproductive isolation between ecotypes of Panicum hallii. *New Phytologist*, *205*, 402–414.

Lowry, D. B., Purmal, C. T., & Juenger, T. E. (2013). A population genetic transect of Panicum hallii (Poaceae). *American Journal of Botany*, *100*(3), 592–601.

Lowry, D. B., Rockwood, R. C., & Willis, J. H. (2008). Ecological reproductive isolation of coast and inland races of Mimulus guttatus. *Evolution*, *62*(9), 2196–2214.

Lowry, D. B., Taylor, S. H., Bonnette, J., Aspinwall, M. J., Asmus, A. L., Keitt, T. H., … Juenger, T. E. (2015). QTLs for Biomass and Developmental Traits in Switchgrass (Panicum virgatum). *BioEnergy Research*.

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., … Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genetics*, *9*(1), e1003215.

Luterbacher, J. S., Rand, J. M., Martin Alonso, D., Han, J., Youngquist, J. T., Maravelias, C. T., … Dumesic, J. A. (2014). Nonenzymatic Sugar Production from Biomass Using Biomass-Derived g -Valerolactone. *Science*, *343*(January), 277–281.

Lynch, M., & Hill, W. G. (1986). Phenotypic Evolution by Neutral Mutation. *Evolution*, *40*(5), 915–935.

Mackay, T. F. C., Stone, E. a, & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature reviews. Genetics*, *10*(8), 565–577.

Manichaikul, A., Moon, J. Y., Sen, Ś., Yandell, B. S., & Broman, K. W. (2009). A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics*, *181*(3), 1077–1086.

Maranan, M. C., & Laborie, M. P. G. (2008). Rapid prediction of the chemical traits of hybrid poplar with near infrared spectroscopy. *Journal of Biobased Materials and Bioenergy*, *2*(1), 57–63.

Markwell, J., Osterman, J. C., & Mitchell, J. L. (1995). Calibration of the Minolta SPAD-502 leaf chlorophyll meter. *Photosynthesis Research*, *46*(3), 467–472.

Martin, G., Chapuis, E., & Goudet, J. (2008). Multivariate Qst-Fst comparisons: a neutrality test for the evolution of the G matrix in structured populations. *Genetics*, *108*, 2135–2149.

Martínez-Reyna, J. M., & Vogel, K. P. (2002). Crop Breeding, Genetics & Cytology: Incompatibility Systems in Switchgrass. *Crop Science*, *42*, 1800–1805.

McKay, J. K., & Latta, R. G. (2002). Adaptive population divergence: Markers, QTL and traits. *Trends in Ecology and Evolution*.

McLaughlin, S. B., & Kszos, L. A. (2005). Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*, *28*(6), 515–535.

McMillan, C. (1959). The role of ecotypic variation in the distribution of the central grassland of North America. *Ecological Monographs*, *29*(4), 286–308.

McMillan, C. (1964). Ecotypic Differentiation Within Four North American Prairie Grasses. I. Morphological Variation Within Transplanted Community Fractions. *American Journal of Botany*, *51*(10), 1119–1128.

McMillan, C. (1965). Ecotypic Differentiation Within Four North American Prairie Grasses. II. Behavioral Variation Within Transplanted Community Fractions. *American Journal of Botany*, *52*(1), 55–65.

McMillan, C. (1967). Phenological Variation within Seven Transplanted Grassland Community Fractions from Texas and New Mexico. *Ecology*, *48*(5), 807–813.

Melendez-Ackerman, E. J. (1997). Patterns of Color and Nectar Variation Across an Ipomopsis (Polemoniaceae) Hybrid Zone. *American Journal of Botany*.

Melendez-Ackerman, E. J., & Campbell, D. R. (1998). Adaptive significance of flower color and inter-trait correlations in an Ipomopsis hybrid zone. *Evolution*, *52*(5), 1293–1303.

Merilä, J., & Crnokrak, P. (2001). Comparison of genetic differntiation at marker loci and quanitative traits. *Journal of Evolutionary Biology*, *14*, 892–903.

Meyer, E., Logan, T. L., & Juenger, T. E. (2012). Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research. *Plant Journal*, *70*(5), 879–890.

Mitchell, R. J. (1993). Adaptive significance of Ipomopsis aggregata nectar production: observation and experiment in the field. *Evolution*, *47*(1), 25–35.

Mitchell, R. J. (1994). Effects of Floral Traits, Pollinator Visitation, and Plant Size on Ipomopsis aggregata Fruit Production. *The American Naturalist*.

Mitchell, R. J. (2004). Heritability of nectar traits: why do we know so little? *Ecology*, *85*(6), 1527–1533.

Mitchell, R. J., & Waser, N. M. (1992). Adaptive significance of Ipomopsis aggregata nectar production: pollination success of single flowers. *Ecology*.

Moles, A. T., Peco, B., Wallis, I. R., Foley, W. J., Poore, A. G. B., Seabloom, E. W., … Hui, F. K. C. (2013). Correlations between physical and chemical defences in plants: Tradeoffs, syndromes, or just many different ways to skin a herbivorous cat? *New Phytologist*, *198*(1), 252–263.

Monti, A., Di Virgilio, N., & Venturi, G. (2008). Mineral composition and ash content of six major energy crops. *Biomass and Bioenergy*, *32*(3), 216–223.

Morris, G. P., Grabowski, P. P., & Borevitz, J. O. (2011). Genomic diversity in switchgrass (Panicum virgatum): From the continental scale to a dune landscape. *Molecular Ecology*, *20*(23), 4938–4952.

Nabity, P. D., Orpet, R., Miresmailli, S., Berenbaum, M. R., & DeLucia, E. H. (2012). Silica and nitrogen modulate physical defense against chewing insect herbivores in bioenergy crops Miscanthus x Giganteus and Panicum virgatum (Poaceae). *J Econ Entomol*, *105*, 878–883.

Nelson, C. J., & Moser, L. E. (1994). Plant Factors Affecting Forage Quality. In G. C. Fahey (Ed.), *Forage Quality, Evaluation, and Utilization* (pp. 115–154). Lincoln, NE: Anerican Society of Agronomy, Inc.

Nuismer, S. L., Thompson, J. N., & Gomulkiewicz, R. (1999). Gene flow and geographically structured coevolution. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1419), 605.

O'hara, R. B., & Merilä, J. (2005). Bias and Precision in QST Estimates: Problems and Some Solutions. *Genetics*, *171*(3), 1331–1339.

Oba, M., & Allen, M. S. (1999). Evaluation of the importance of the digestibility of neutral detergent fiber from forage: effects on dry matter intake and milk yield of dairy cows. *Journal of dairy science*, *82*(3), 589–596.

Okada, M., Lanzatella, C., Saha, M. C., Bouton, J., Wu, R., & Tobias, C. M. (2010). Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics*, *185*(3), 745–760.

Orr, H. A. (1998). The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution*, *52*(4), 935–949.

Ovaskainen, O., Karhunen, M., Zheng, C., Arias, J. M. C., & Merilä, J. (2011). A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics*, *189*(2), 621–632.

Parrish, D., Casler, M., & Monti, A. (2012). The Evolution of Switchgrass as an Energy Crop. In A. Monti (Ed.), *Switchgrass* (pp. 1–28). London: Springer.

Patakas, A. (2012). Abiotic Stress-Induced Morphological and Anatomical Changes in Plants. In P. Ahmad & M. N. V. Prasad (Eds.), *Abiotic Stress Responses in Plants: Metabolism, Productivity and Sustainability* (pp. 267–281). Springer Science+Business Media, LLC.

Pawar, P. M.-A., Koutaniemi, S., Tenkanen, M., & Mellerowicz, E. J. (2013). Acetylation of woody lignocellulose: significance and regulation. *Frontiers in plant science*, *4*(May), 118.

Payne, C. E., & Wolfrum, E. J. (2015). Rapid analysis of composition and reactivity in cellulosic biomass feedstocks with near-infrared spectroscopy. *Biotechnology for Biofuels*, *8*(1), 1–14.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., … Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, *61*, 167–234.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, *7*(5), e37135–e37135.

Porter, C. L. J. (1966). An Analysis of Variation Between Upland and Lowland Switchgrass, Panicum Virgatum L., in Central Oklahoma. *Ecology*, *47*(6), 980–992.

Porter, J. M., & Johnson, L. A. (2000). A phylogenetic classification of Polemoniaceae. *Aliso*, *19*, 55–91.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Provan, J., Powell, W., & Hollingsworth, P. M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in ecology & evolution*, *16*(3), 142–147.

Rajamma, R., Ball, R. J., Tarelho, L. a C., Allen, G. C., Labrincha, J. a., & Ferreira, V. M. (2009). Characterisation and use of biomass fly ash in cement-based materials. *Journal of Hazardous Materials*, *172*(2-3), 1049–1060.

Ramsey, J., Bradshaw Jr, H., & Schemske, D. (2003). Components of reproductive isolation between the monkeyflowers Mimulus lewisii and M. cardinalis (Phrymaceae). *Evolution*, *57*(7), 1520–1534.

Rausher, M. D. (2008). Evolutionary transitions in floral color. *International Journal of Plant Sciences*, *169*(1), 7–21.

Reusch, T. B. H., & Wood, T. E. (2007). Molecular ecology of global change. *Molecular ecology*, *16*(19), 3973–3992.

Roberts, C. A., Houx, J. H., & Fritschi, F. B. (2011). Near-infrared analysis of sweet sorghum bagasse. *Crop Science*, *51*(5), 2284–2288.

Roberts, C. A., Workman, J. J., & Reeves, J. B. I. (2004). *Near-Infrared Spectroscopy in Agriculture*. (C. A. Roberts, J. Workman Jr., & J. B. Reeves III, Eds.). Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Rogers, A. R., & Harpending, H. C. (1983). Population structure and quantitative characters. *Genetics*, *105*(4), 985–1002.

Sangster, A. G., & Hodson, M. J. (1986). Silica in higher plants. *Ciba Foundation symposium*, *121*, 90–107.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature reviews. Genetics*, *14*(11), 807–20.

Schiemenz, K., & Eichler-Löbermann, B. (2010). Biomass ashes and their phosphorus fertilizing effect on different crops. *Nutrient Cycling in Agroecosystems*, *87*(3), 471–482.

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, *17*(1), 30–44.

Serba, D. D., Daverdin, G., Bouton, J. H., Devos, K. M., Brummer, C. E., & Saha, M. C. (2014). Quantitative Trait Loci (QTL) Underlying Biomass Yield and Plant Height in Switchgrass. *BioEnergy Research*, *8*(1), 307–324.

Serba, D., Wu, L., Daverdin, G., Bahri, B. a., Wang, X., Kilian, A., … Devos, K. M. (2013). Linkage Maps of Lowland and Upland Tetraploid Switchgrass Ecotypes. *Bioenergy Research*, *6*(3), 953–965.

Sinervo, B., & Svensson, E. (2002). Correlational selection and the evolution of genomic architecture. *Heredity*, *89*, 329–338.

Sluiter, J. B., Ruiz, R. O., Scarlata, C. J., Sluiter, A. D., & Templeton, D. W. (2010). Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. *Journal of Agricultural and Food Chemistry*, *58*(16), 9043–9053.

Somerville, C., Youngs, H., Taylor, C., Davis, S. C., & Long, S. P. (2010). Feedstocks for Lignocellulosic Biofuels. *Science*, *329*, 790–793.

Spitze, K. (1993). Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. *Genetics*, *135*(2), 367–374.

Stearns, F., Boles, S., Hurston, H., Vo, T., Butler, D., Shuham, W., & Juenger, T. E. (2008). Identification of nuclear microsatellite loci for Ipomopsis aggregata and the distribution of pairwise relatedness in a natural population. *Molecular Ecology Resources*, *8*(2), 437–439.

Stebbins, G. L. (1970). Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms. *Annual Review of Ecology and Systematics*, *1*, 307–326.

Streisfeld, M. a, & Kohn, J. R. (2005). Contrasting patterns of floral and molecular variation across a cline in Mimulus aurantiacus. *Evolution; international journal of organic evolution*, *59*(12), 2548–2559.

Sun, F., Flanders, W., Yang, Q. H., & Zhao, H. Y. (2000). Transmission/disequilibrium tests for quantitative traits. *Annals of Human Genetics*, *64*, 555–565.

Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. *Nature reviews. Genetics*, *9*(6), 444–457.

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular Ecology*, *15*(2), 343–355.

Theander, O., Aman, P., Westerlund, E., Andersson, R., & Pettersson, D. (1995). Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason lignin (the Uppsala method): collaborative study. *Journal of AOAC International*, *78*(4), 1030–1044.

Thompson, J. N. (1994). *The Coevolutionary Process*. Chicago: The University of Chicago Press.

Thompson, J. N. (1999). Specific Hypotheses on the Geographic Mosaic of Coevolution. *The American Naturalist*, *153*(s5), S1–S14.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, *29*(12), 673–680.

Tilman, D., Hill, J., & Lehman, C. (2006). Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. *Science*, *314*, 1598–1601.

Tripathi, D. K., Singh, V. P., Gangwar, S., Prasad, S. M., Maurya, J. N., & Chauhan, D. K. (2014). Role of Silicon in Enrichment of Plant Nutrients and Protection from Biotic and Abiotic Stresses. In P. Ahmad, M. R. Wani, M. M. Azooz, & L.-S. P. Tran (Eds.), *Improvement of Crops in the Era of Climatic Changes* (pp. 153–190). Springer New York.

Udén, P., Robinson, P. H., & Wiseman, J. (2005). Use of detergent system terminology and criteria for submission of manuscripts on new, or revised, analytical methods as well as descriptive information on feed analysis and/or variability. *Animal Feed Science and Technology*, *118*(3-4), 181–186.

Uppalapati, S. R., Serba, D. D., Ishiga, Y., Szabo, L. J., Mittal, S., Bhandari, H. S., … Saha, M. C. (2013). Characterization of the Rust Fungus, Puccinia emaculata, and Evaluation of Genetic Variability for Rust Resistance in Switchgrass Populations. *Bioenergy Research*, *6*(2), 458–468.

Van Ooijen, J. W. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics research*, *93*(5), 343–349.

Van Soest, P. J., & Wine, R. H. (1967). Use of detergent in the analysis of fibrous feeds. IV. Determination of plant cell wall constituents. *J. Assoc. Off. Ana. Chem.*, *50*(5), 50–55.

Van Zyl, W. H., & den Haan, R. (2013). Developing Cellulolytic Organisms for Consolidated Bioprocessing of Lignocellulosics. In V. K. Gupta & M. G. Tuohy (Eds.), *Biofuel Technologies* (pp. 189–220). Springer Berlin Heidelberg.

Vermerris, W., Saballos, A., Ejeta, G., Mosier, N. S., Ladisch, M. R., & Carpita, N. C. (2007). Molecular breeding to enhance ethanol production from corn and sorghum stover. In *Crop Science* (Vol. 47).

Vogel, K. P., Dien, B. S., Jung, H. G., Casler, M. D., Masterson, S. D., & Mitchell, R. B. (2011). Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *Bioenergy Research*, *4*(2), 96–110.

Vogel, K. P., & Jung, H.-J. G. (2001). Genetic Modification of Herbaceous Plants for Feed and Fuel. *Critical Reviews in Plant Sciences*, *20*(1), 15–49.

Wade, M. J., & Goodnight, C. J. (1998). Perspective: The Theories of Fisher and Wright in the Context of Metapopulations: When Nature Does Many Small Experiments. *Evolution*, *52*(6), 1537–1553.

Walsh, B., & Blows, M. W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 41–59.

Wang, Y., Shao, Y., Matovic, M. D., & Whalen, J. K. (2014). Recycling of switchgrass combustion ash in cement: Characteristics and pozzolanic activity with chemical accelerators. *Construction and Building Materials*, *73*, 472–478.

Waser, N., Chittka, L., Price, M., Williams, N. M., & Ollerton, J. (1996). Generalization in pollination systems, and why it matters. *Ecology*, *7*(4), 1043–1060.

Whitlock, M. C. (2008). Evolutionary inference from Q ST. *Molecular Ecology*, *17*(8), 1885–1896.

Whitlock, M. C., & Guillaume, F. (2009). Testing for Spatially Divergent Selection: Comparing QST to FST. *Genetics*, *183*(3), 1055–1063.

Wilken, D., & Allard, S. (1986). Intergradation among Populations of the Ipomopsis aggregata Complex in the Colorado Front Range. *Systematic Botany*, *11*(1), 1–13.

Wolf, P. G., & Soltis, P. S. (1992). Estimates of gene flow among populations, geographic races, and species in the Ipomopsis aggregata complex. *Genetics*, *130*(3), 639–647.

Wolfrum, E. J., Lorenz, A. J., & deLeon, N. (2009). Correlating detergent fiber analysis and dietary fiber analysis data for corn stover collected by NIRS. *Cellulose*, *16*(4), 577–585.

Wolfrum, E. J., Payne, C., Stefaniak, T., Rooney, W., Dighe, N., Bean, B., & Dahlberg, J. (2013). *Multivariate Calibration Models for Sorghum Composition using Near-Infrared Spectroscopy Multivariate Calibration Models for Sorghum Composition using Near-Infrared Spectroscopy*. Golden, CO.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Wright, S. (1951). The Genetical Structure of Populations. *Annals of Eugenics*, *15*, 323–354.

Wu, S., Yang, J., Huang, Y., Li, Y., Yin, T., Wullschleger, S. D., … Wu, R. (2010). An improved approach for mapping quantitative trait Loci in a pseudo-testcross: revisiting a poplar mapping study. *Bioinform Biol Insights*, *4*, 1–8.

Wullschleger, S. D., Sanderson, M. a., McLaughlin, S. B., Biradar, D. P., & Rayburn, a. L. (1996). Photosynthetic rates and ploidy levels among populations of switchgrass. *Crop Science*, *36*(2), 306–312.

Yang, J., Worley, E., Wang, M., Lahner, B., Salt, D. E., Saha, M., & Udvardi, M. (2009). Natural variation for nutrient use and remobilization efficiencies in switchgrass. *Bioenergy Research*, *2*(4), 257–266.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*(7), 1897–1911.

Zhang, S.-J., Song, X.-Q., Yu, B.-S., Zhang, B.-C., Sun, C.-Q., Knox, J. P., & Zhou, Y.-H. (2012). Identification of quantitative trait loci affecting hemicellulose characteristics based on cell wall composition in a wild and cultivated rice species. *Molecular plant*, *5*(1), 162–75.

Zhang, Y., Zalapa, J. E., Jakubowski, A. R., Price, D. L., Acharya, A., Wei, Y., … Casler, M. D. (2011a). Natural hybrids and gene flow between upland and lowland switchgrass. *Crop Science*, *51*(6), 2626–2641.

Zhang, Y., Zalapa, J. E., Jakubowski, A. R., Price, D. L., Acharya, A., Wei, Y., … Casler, M. D. (2011b). Post-glacial evolution of Panicum virgatum: Centers of diversity and gene pools revealed by SSR markers and cpDNA sequences. *Genetica*, *139*(7), 933–948.